



**Kauno technologijos universitetas**

Informatikos fakultetas

**Prekių ženklų reklamos efektyvumo vertinimas taikant giliojo mokymosi algoritmus analizuojant vaizdinę medžiagą**

Baigiamasis magistro projektas

---

**Julius Gudauskas**

Projekto autorius

**Prof. Agnė Paulauskaitė - Tarasevičienė**

Darbo vadovė

---

**Kaunas, 2024**



**Kauno technologijos universitetas**

Informatikos fakultetas

**Prekių ženklų reklamos efektyvumo vertinimas taikant giliojo mokymosi algoritmus analizuojant vaizdinę medžiagą**

Baigiamasis magistro projektas

Dirbtinio intelekto informatika 6211BX007

---

**Julius Gudauskas**

Projekto autorius

**Prof. Agnė Paulauskaitė - Tarasevičienė**

Darbo vadovė

**Prof. Alfonsas Misevičius**

Recenzentas

---

**Kaunas, 2024**



**Kauno technologijos universitetas**

Informatikos fakultetas

Julius Gudauskas

## **Prekių ženklų reklamos efektyvumo vertinimas taikant giliojo mokymosi algoritmus analizuojant vaizdinę medžiagą**

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autoriaus ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Julius Gudauskas



**Kauno technologijos universitetas**

Informatikos fakultetas

## **Baigiamojo magistro projekto užduotis**

Projekto tema

Prekių ženklų reklamos efektyvumo vertinimas taikant giliojo mokymosi algoritmus analizuojant vaizdinę medžiagą

---

Reikalavimai ir sąlygos  
(tikslinti pavadinimą  
pagal poreikį)

Vadovas / Vadovė

Prof. Agnė Paulauskaitė - Tarasevičienė

---

(vadovo pareigos, vardas, pavardė, parašas)

(data)

Julius Gudauskas. Prekių ženklų reklamos efektyvumo vertinimas taikant giliojo mokymosi algoritmus analizuojant vaizdinę medžiagą. Magistro baigiamasis projektas / vadovė prof. Agnė Paulauskaitė - Tarasevičienė; Kauno technologijos universitetas, Informatikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Dirbtinio intelekto informatika.

Reikšminiai žodžiai: gilusis mokymasis, dirbtinis intelektas, objektų aptikimas, reklamos matomumas.

Kaunas, 2024. 75 p.

### **Santrauka**

Dinamiškoje rinkodaros ir reklamos aplinkoje, siekiant suprasti prekės ženklų eksponavimo vietų poveikį, prekės ženklų reklamų matomumo vertinimas atlieka itin svarbų vaidmenį. Tradiciniai, rankinio anotavimo ir analizės metodai, pareikalauja daug laiko ir yra subjektyvūs, todėl efektyviam ir objektyviam vertinimui reikalingi automatizuoti sprendimai. Nors šie sprendimai yra labai aktualūs, tyrimų, kuriuose būtų sprendžiama atvira tiek žinomų, tiek nežinomų prekių ženklų reklamos paieška, yra nedaug. Šiame tyrime, panaudojant giliojo mokymosi algoritmus, pasiūlomas naujas metodas, kuris geba analizuoti tiesioginių sporto transliacijų vaizdo įrašus ir pateikti prekės ženklų matomumo rezultatus. Modelių mokymui panaudoti atviri logotipų duomenų rinkiniai, taip pat sukurtas ir panaudotas naujas krepšinio rungtynių vaizdų duomenų rinkinys su anotuotomis prekių ženklų reklamomis. Eksperimentuose aprašomas modelių veikimas ir našumas.

Julius Gudauskas. Evaluating brand advertising effectiveness using deep learning algorithms in visual material analysis. Master's Final Degree Project / supervisor prof. Agnė Paulauskaitė - Tarasevičienė; Faculty of Informatics, Kaunas University of Technology.

Study field and area (study field group): Artificial Intelligence in Computer Science

Keywords: deep learning, artificial intelligence, object detection, brands visibility.

Kaunas, 2024. 75 p.

### **Summary**

In the dynamic landscape of marketing and advertising, assessing brand visibility in live sports events plays a pivotal role in understanding brand exposure and impact. Traditional methods of manual annotation and analysis are time-consuming and subjective, necessitating automated solutions for efficient and objective evaluation. Although these decisions are highly relevant, there is a small amount of research that addresses open-ended advertising detection for both known and unknown brands. This study proposes a new method using a deep learning algorithm that can analyze live sports videos and provide brands visibility results. Open logo datasets were utilized for model training, along with a newly created and annotated dataset of brand advertisements from basketball game images. The experiments conducted provide detailed insights into the models' behavior and performance.

## Turinys

Lentelių sąrašas .....	9
Santrumpų ir terminų sąrašas .....	12
Įvadas .....	14
1. Reklamos efektyvumo vertinimui pritaikomų įrankių analizė.....	16
1.1. Reklamos efektyvumo vertinimo svarbos bei paklausos analizė.....	16
1.2. Atviri logotipų duomenų rinkiniai .....	17
1.2.1. FlickrLogos su 27 klasėmis.....	17
1.2.2. FlickrLogos su 32 klasėmis.....	17
1.2.3. BelgaLogos .....	18
1.2.4. Logo-Net .....	18
1.2.5. LogoDet-3K .....	18
1.3. Giliuoju mokymusi grįstos objektų aptikimo ir atpažinimo architektūros .....	18
1.3.1. Konvoliuciniai neuroniniai tinklai .....	18
1.3.1.1. Konvoliucinių neuroninių tinklų mokymo procesas .....	20
1.3.1.2. Konvoliucinių neuroninių tinklų trūkumai .....	20
1.3.2. ResNet – liekamasis neuroninis tinklas .....	21
1.3.2.1. Skirtingų sluoksnių ResNet.....	21
1.3.3. DenseNet.....	22
1.3.4. VGGNet .....	22
1.3.5. Regionais grįsti konvoliuciniai neuroniniai tinklai .....	23
1.3.5.1. R-CNN, greitas R-CNN ir greitesnis R-CNN.....	23
1.3.6. YOLO.....	24
1.3.6.1. YOLO architektūra .....	25
1.4. Logotipų aptikimo ir atpažinimo sprendimų analizė .....	25
1.4.1. Logotipų atpažinimas naudojant ResNet .....	25
1.4.2. Logotipų atpažinimas naudojant DenseNet .....	26
1.4.3. Greito R-CNN ir greitesnio R-CNN taikymas prekių ženklų aptikime .....	27
1.4.4. Logotipų aptikimas naudojant YOLO.....	27
1.5. Objektų aptikimo ir atpažinimo sprendimų analizė .....	28
1.5.1. ResNet ir YOLO hibridinė sistema .....	28
1.5.2. Objektų aptikimas naudojant DenseNet.....	29
1.6. Giliuoju mokymosi grįstas panašių objektų grupavimas .....	30
1.7. Logotipų aptikimo problemos .....	31
1.7.1. Mažų logotipų aptikimo problema .....	31
1.7.2. Logotipų matomumo kompleksiskumas .....	32
1.7.3. Duomenų rinkinių disbalansas .....	33
1.7.4. Atviro rinkinio logotipų aptikimo problema .....	34
2. Reklamos efektyvumo vertinimo architektūra .....	35
2.1. Sistemos panaudojimo atvejai.....	35
2.2. Sistemos statinis vaizdas .....	36
2.3. Funkciniai ir nefunkciniai reikalavimai .....	36
2.4. Sistemos veikimo detalizavimas .....	37
2.4.1. Bendras kuriamos sistemos aprašas .....	37
2.4.2. Vaizdo įrašo apdorojimo panaudos atvejo aprašas .....	40

2.4.3.	Reklamos efektyvumo vertinimo duomenų apdorojimo panaudos atvejo aprašas .....	41
3.	Atviri bei sudaryti prekių ženklų duomenų rinkiniai .....	42
3.1.	Atvirų duomenų rinkinių aprašymai ir jų analizė .....	42
3.2.	Duomenų rinkinio kūrimas .....	44
3.3.	Duomenų rinkinių apdorojimas .....	46
4.	Reklamos efektyvumo vertinimas .....	49
5.	Prekių ženklų aptikimo bei atpažinimo tiriamoji dalis .....	51
5.1.	Tiriamosios dalies aprašymas .....	51
5.2.	Logotipų aptikimo rezultatai .....	55
5.3.	Prekių ženklų reklamų aptikimo tikslumo rezultatai krepšinio rungtynių kadruose .....	62
5.4.	Prekių ženklų atpažinimo bei klasifikavimo testavimo rezultatai .....	66
6.	Prekių ženklų reklamos efektyvumo vertinimo sistemos naudotojo vadovas .....	68
	Išvados .....	71
	Literatūros sąrašas .....	72
	Priedai .....	75



## Lentelių sąrašas

<b>1 lentelė.</b> Atvirų duomenų rinkinių aprašymas.....	17
<b>2 lentelė.</b> Skirtingų <i>ResNet</i> palyginimas.....	21
<b>3 lentelė.</b> <i>ResNet50</i> ir autoriaus pasiūlyto modelio palyginimas.....	26
<b>4 lentelė.</b> Skirtingų testuotų modelių rezultatų palyginimas .....	29
<b>5 lentelė.</b> „Atlikti vaizdo įrašo analizę“ panaudos atvejo aprašas .....	35
<b>6 lentelė.</b> „Apdoroti reklamos efektyvumo vertinimo duomenis“ panaudos atvejo aprašas .....	35
<b>7 lentelė.</b> FlickrLogos duomenų rinkinio aprašas .....	43
<b>8 lentelė.</b> <i>LogoDet-3k</i> duomenų rinkinio aprašas .....	44
<b>9 lentelė.</b> Duomenų rinkinių klasių pasiskirstymas po apdorojimo.....	47
<b>10 lentelė.</b> YOLOv7 rezultatai su FlickrLogos-32 .....	60
<b>11 lentelė.</b> Greitesnio R-CNN rezultatai su FlickrLogos-32 .....	60
<b>12 lentelė.</b> YOLOv7 rezultatai su <i>LogoDet-3K</i> .....	60
<b>13 lentelė.</b> Greitesnio R-CNN rezultatai su <i>LogoDet-3K</i> .....	60
<b>14 lentelė.</b> Prekių ženklų grupavimui taikytų vaizdo savybių išgavimo architektūrų bendro tikslumo rezultatai.....	66

## Paveikslėlių sąrašas

<b>1 pav.</b> Bendra CNN architektūra .....	18
<b>2 pav.</b> Telkimo sluoksnio veikimo vizualizacija .....	20
<b>3 pav.</b> FFNN ir <i>ResNet</i> palyginimas .....	21
<b>4 pav.</b> <i>DenseNet</i> tinklo architektūra .....	22
<b>5 pav.</b> R-CNN veikimo principo diagrama.....	23
<b>6 pav.</b> Greito R-CNN veikimo principo diagrama .....	24
<b>7 pav.</b> Greitesnio R-CNN veikimo principo daigramas.....	24
<b>8 pav.</b> YOLO iliustruotas veikimo principas [27] .....	24
<b>9 pav.</b> YOLO architektūra .....	25
<b>10 pav.</b> Autoriaus pasiūlyto modelio architektūra.....	26
<b>11 pav.</b> <i>DenseNet</i> interpretacija logotipų atpažinimo uždaviniui spręsti.....	27
<b>12 pav.</b> <i>ResNet</i> ir YOLO hibridinė architektūra .....	28
<b>13 pav.</b> DarkNet-53 architektūra .....	29
<b>14 pav.</b> <i>ResNet</i> ir <i>DenseNet</i> rezultatų palyginimas [20].....	30
<b>15 pav.</b> DeepCluster architektūra [38].....	31
<b>16 pav.</b> Skirtingų oro sąlygų duomenų pavyzdžiai.....	33
<b>17 pav.</b> Panaudos atvejų diagrama .....	35
<b>18 pav.</b> Vaizdo įrašų analizės sistemos klasių diagrama .....	36
<b>19 pav.</b> Darbe sukurtų mašininio mokymosi dalių tarpusavio komunikacija .....	37
<b>20 pav.</b> Sukurtos sistemos veikimo principo schema .....	39
<b>21 pav.</b> „Atlikti vaizdo įrašo analizę“ veiklos diagrama .....	40
<b>22 pav.</b> „Apdoroti reklamos efektyvumo vertinimo duomenis“ veiklos diagrama .....	41
<b>23 pav.</b> <i>FlickrLogos</i> - 32 duomenų peržiūra.....	42
<b>24 pav.</b> <i>LogoDet-3k</i> duomenų peržiūra .....	42
<b>25 pav.</b> <i>SportLogo</i> duomenų peržiūra.....	43
<b>26 pav.</b> <i>LogoDet-3K</i> duomenų rinkinio analizė.....	44
<b>27 pav.</b> Sudaryto duomenų rinkinio peržiūra .....	45
<b>28 pav.</b> Sukurto prekės ženklų krepšinio rungtyne duomenų rinkinio analizė .....	46
<b>29 pav.</b> Duomenų rinkinio apdorojimo eiga .....	47
<b>30 pav.</b> Prekių ženklų užimamų plotų vizualizavimas .....	49
<b>31 pav.</b> Prekių ženklų atstumo iki centro vizualizavimas .....	50
<b>32 pav.</b> Kelių tų pačių prekių ženklų akcentavimas .....	50
<b>33 pav.</b> YOLOv7 architektūros vizualizavimas [21].....	51
<b>34 pav.</b> Greitesnio R-CNN architektūros vizualizavimas [22].....	53
<b>35 pav.</b> Skirtingų FPN rezultatų palyginimas.....	55
<b>36 pav.</b> <i>ResNet50 v2</i> nuostolių verčių kaita mokymo metu .....	56
<b>37 pav.</b> YOLOv7 nuostolių verčių kaita mokymo metu .....	56
<b>38 pav.</b> YOLOv7 tikslumo verčių kaita mokymo metu .....	57
<b>39 pav.</b> Greitesnio R-CNN ir YOLOv7 tikslumo verčių palyginimas .....	58
<b>40 pav.</b> YOLOv7 klaidų matrica .....	58
<b>41 pav.</b> Logotipų duomenų rinkinio anotacijos (viršuje) bei prognozės (apačioje) .....	59
<b>42 pav.</b> Keblūs aptikimo atvejai .....	61
<b>43 pav.</b> Krepšinio rungtynių anotacijos (viršuje) bei prognozės (apačioje) .....	62
<b>44 pav.</b> YOLOv7 nuostoliai mokymo metu .....	63

<b>45 pav.</b> YOLOv7 tikslumo verčių kaita mokymo metu .....	63
<b>46 pav.</b> YOLO bendro tikslumo su skirtingais IOU slenksčiais reikšmių kaita mokymo metu.....	64
<b>47 pav.</b> YOLOv7 klaidų matrica .....	64
<b>48 pav.</b> Krepšinio rungtynių anotacijos (viršuje) bei aptikimo rezultatai po papildomo mokymo (apačioje).....	65
<b>49 pav.</b> Prekių ženklų suskirstymo pagal vaizdines savybes eksperimento rezultatai .....	66
<b>50 pav.</b> Prekių ženklų suskirstymo rezultatų vizualizavimas .....	67
<b>51 pav.</b> Rezultatų failo pavyzdys.....	68
<b>52 pav.</b> Prekių ženklų reklamų aptikimas ir sekimas .....	69
<b>53 pav.</b> Aptikti prekės ženklai išsaugoti aplanke .....	69
<b>54 pav.</b> Prekių ženklų grupės ir grupės rezultatai.....	70

## Santrumpų ir terminų sąrašas

### Santrumpos:

**ROI** – investicijos grąža (angl. *return of investment*)

**SVM** – rėmėjo žiniasklaidos vertė (angl. *sponsor media value*)

**FVM** – pilna žiniasklaidos vertė (angl. *full media value*)

**MVP** – medijos vertės dalis (angl. *media value percentage*)

**EMV** – uždirbta medijos vertė (angl. *earned media value*)

**CNN** – konvoliucinis neuroninis tinklas (angl. *convolutional neural network*)

**reLU** – ištaisytas tiesinis vienetas (angl. *rectified linear unit*)

**FC** – pilnai sujungtas (angl. *fully connected*)

**FFNN** – grįžtamasis neuroninis tinklas (angl. *feedforward neural network*)

**VGG** – vizualinės geometrijos grupė (angl. *visual geometry group*)

**RPN** – regionų pasiūlymo (angl. *region proposal network*)

**mAP** – vidutinis tikslumas (angl. *mean average precision*)

**IoU** - sankryža virš sąjungos (angl. *intersection over union*)

**CAL** – kontekstinis priešpriešinasis mokymasis (angl. *contextual adversarial learning*)

**EDA** – į įvykius orientuota architektūra (angl. *event driven architecture*)

**CRAFT** – tekstinių simbolių regionų suvokimas (angl. *character region awareness for text*)

**COCO** – įprasti objektai kontekste (angl. *common objects in context*)

**SIFT** - nekintamų mastelio savybių transformacija (angl. *scale-invariant feature transform*)

**HOG** - orientuotų gradientų histograma (angl. *histogram of oriented gradients*)

**ML** – mašininis mokymasis (angl. *machine learning*)

**SGD** - stochastinis gradiento nusileidimas (angl. *stochastic gradient descent*)

**HSV** – atspalvis, sodrumas ir vertė (angl. *hue saturation value*)

### Terminai:

**Neuroninis tinklas** – kompiuterinis modelis, įkvėptas biologinių smegenų struktūros, naudojamas mokymuisi iš duomenų. Jį sudaro neuronai (mazgai) ir jungtys (svoriai).

**Pasikartojantis neuroninis tinklas (RNN)** – yra neuroninio tinklo tipas, kai ankstesnio veiksmo išvestis įvedama į dabartinį veiksmą.

**Konvoliucija** – matematinė operacija, leidžianti sujungti du informacijos rinkinius į vieną naują.

**Konvergencija** – taškas, kai mokymo procesas pasiekia stabilią būseną, o tinklo parametrai (t. y. svoriai ir poslinkiai) nusistovi ties reikšmėmis, kurios sukuria tikslias mokymo duomenų prognozes.

**Klasteris** – panašių objektų grupė

**Koreliacija** – statistinis ryšys tarp kintamųjų

**Aktyvacijos funkcija** – funkcija, kuri apskaičiuoja mazgo išvestį pagal atskiras jo įvestis ir jų svorius.

## Ivadas

### Problemos aktualumas

Vaizdo įrašuose bei tiesioginėse transliacijose galime pastebėti gausybę prekių ženklų. Dalyje vaizdo įrašų prekės ženklo reklamos užsakovams yra užtikrinamos susitartos prekės ženklo matomumo sąlygos – koku formatu, kiek laiko, kokiomis sąlygomis jis bus matomas. Tačiau tam tikrais atvejais šių sąlygų užtikrinti nepavyksta – prekės ženklo matomumas ir efektyvumas išlieka nežinomas. Rėmimo sutartys dažnai sudaromos už didelę kainą, todėl prekių ženklų savininkai yra suinteresuoti sužinoti, kaip efektyviai jų išlaidos buvo panaudotos. Tai ypač aktualu tiesioginių transliacijų vaizdo įrašuose – remiantis „Eventmarketer“ [1] tyrimu, 72 % reklamą stebėjusių žiūrovų išreiškė teigiamą požiūrį į reklamas, kai jos matomos patiriant teigiamas emocijas – stebint muzikos festivalį, sportines varžybas ir pan. Tokie renginiai suburia dideles auditorijas žiūrovų ir tai suteikia prekių ženklams puikias sąlygas pritraukti naujų vartotojų. Tačiau prekės ženklas turi būti patalpintas matomoje vietoje. Tyrimai parodė, kad tokios vietos, kaip ribinės arenos linijos, sienelės, laikomos geriausiomis vietomis rodyti prekės ženklo logotipus, nes tai neerzina žiūrovų ir užtikrina maksimalų matomumą [2] [3].

Šiame darbe analizės objektu pasirinkti tiesioginių krepšinio rungtynių transliacijų vaizdo įrašai. Šio tipo vaizdo įrašuose didžioji dalis prekių ženklų reklamos atributų yra statiniai (t. y. lipdukai, afišos, plakatai), todėl kiek kartų jie buvo matomi vaizdo įrašo transliacijos metu nuspėti ir apskaičiuoti sunku, o įvertinti matomumą reikia nemažai rankinio darbo peržiūrint ir rankiniu būdu skaičiuojant, kiek kartų kuris prekės ženklas pasirodė ekrane. Tačiau tam tikrų matomumo metrikų rankiniu būdu apskaičiuoti nepavyksta, todėl gaunami rezultatai nėra labai išsamūs – jų negalima pilnai panaudoti skirtingų vaizdo įrašų prekių ženklų matomumo palyginime.

Prekės ženklų reklamos efektyvumo vertinimas, panaudojant giliojo mokymosi metodus analizuojant vaizdinę medžiagą, tebėra besivystanti sritis, todėl reikia atlikti tolesnius tyrimus, tobulinti metodikas ir atskleisti įžvalgas, kurios gali padėti priimti strateginius rinkodaros sprendimus. Pagrindiniu iššūkiu išlieka atviras prekių ženklų aptikimo uždavinys – norint apskaičiuoti prekės ženklo matomumo metrikas, reikia sukurti prekių ženklų reklamos aptikimo algoritmą, galintį aptikti visus kadre matomus logotipus bei reklamuojamus objektus. Dauguma egzistuojančių logotipų aptikimo metodų yra sutelkti į ribotą logotipų kategorijų rinkinį, jie geba aptikti tik žinomus prekės ženklus. Šiems modeliams apmokyti panaudojami platūs duomenų rinkiniai, susidedantys iš gausybės skirtingų kategorijų paveikslėlių ir anotacijų [5]. Nepaisant didelio duomenų kiekio, šioje strategijoje pagrindiniu iššūkiu išlieka staigus esamų prekių ženklų kiekio augimas ir spartus dabartinių prekių ženklų įvaizdžių keitimasis. Neretai reklamai naudojami ne tik konkretūs prekių ženklų logotipai, bet ir nauji produktai ar šūkiei, kurie nebūna įtraukti į duomenų rinkinius. Šių objektų tokio tipo modeliai aptikti negali [4].

### Darbo tikslas ir uždaviniai

Šio darbo **tikslas** – taikant giliojo mokymosi algoritmus aptikti vaizdo įrašė matomus prekės ženklus ir pateikti jų reklamos efektyvumo vertinimą.

Tikslui pasiekti išsikelti uždaviniai:

1. išbandyti įvairius giliojo mokymosi vaizdų analizės algoritmus ir įvertinti jų tinkamumą prekių ženklų aptikimo vaizdinėje medžiagoje uždaviniui spręsti.
2. sukurti programinį įrankį, kuris:

- naudodamas giliojo mokymosi algoritmus analizuotų vaizdo medžiagą ir teiktų duomenis apie prekių ženklų matomumą.
- apdorotų ir sistemintų prekių ženklų matomumo duomenis, pateikdamas įvairius efektyvumo rodiklius (reklamos matomumo dažnį, reklamos užimamą plotą ir pan.)

### **Tyrimo objektas**

Šio tyrimo objektas – giliojo mokymosi modelis, kuris aptinka vaizdo įrašuose matomus prekės ženklus bei jų reklamas ir pateikia rezultatus, kuriuos galima interpretuoti vertinant reklamos efektyvumą.

## 1. Reklamos efektyvumo vertinimui pritaikomų įrankių analizė

Šiame skyriuje aprašoma reklamos efektyvumo vertinimo svarba, aptariamos vertinimo metrikos bei egzistuojantys trūkumai. Apžvelgiamos įvairios giliojo mokymosi metodologijos, kurios pritaikomos objektų aptikimo ir atpažinimo uždaviniui spręsti, išskiriami jų pagrindiniai privalumai bei trūkumai. Taip pat aptariami šių metodų praktiniai panaudojimo pavyzdžiai bei literatūroje gauti aptikimo rezultatai.

### 1.1. Reklamos efektyvumo vertinimo svarbos bei paklausos analizė

Prekės ženklo tapatybės ir prekės ženklo žinomumo kūrimas užima svarbią vietą sporto rinkodaros pasaulyje ir yra itin svarbus rėmimo sandoriuose ir partnerystėje. Tyrimai rodo, kad prekės ženklo žinomumas sudaro net 60–80 % [5] bendrų pardavimų, o tai yra labai svarbu ilgalaikiam pajamų augimui. Nenuostabu, kad prekių ženklų rėmimas auga maždaug 4 % per metus – rėmimas jiems yra puikus būdas sustiprinti savo identitetą bei paskatinti žiūrovus domėtis jų prekės ženklu [6].

Norėdami išsiaiškinti, kaip geriausiai nustatyti rėmimo atsiperkamumą ROI, prekės ženklai pirmiausia turi suprasti, kaip įgyti ir išmatuoti matomumą. Aiškus logotipų matomumas bet kuriame sporto renginyje padidina prekės ženklo pardavimą, patraukdamas žiūrinčių gerbėjų, kuriuos natūraliai traukia prekės ženklai, susiję su konkrečia komanda ar sporto šaka, dėmesį [5]. Daugelis prekių ženklų renkasi remti atskirus sportininkus, kurie varžybų metu dėvi jų prekės ženklais puoštą atributiką. Šios atributikos reklamos taip pat matomos ir arenose ar stadionuose. Kai gerbėjai mato, kad jų mėgstamas žaidėjas, kuris dažnai skelbia įrašus įvairiose socialinės žiniasklaidos platformose ir bendrauja su sekėjais, vilki marškinėlius su konkrečia prekės ženklo logotipu, jie yra paskatinami pirkti to prekės ženklo produktus. Be to, stadionuose ir kitose sporto vietose logotipai yra aplink tribūną renginio metu [7].

Tačiau logotipai matomi ne tik renginio metu. John A. Fortunato knygoje „Sporto rėmimas: principai ir praktika“ pažymima, kad sportininkai yra kalbinami prieš ir po treniruotės, o filmuota medžiaga ir nuotraukos iš treniruočių dažnai naudojamos žiniasklaidos pranešimuose [8]. Kitaip tariant, interviu skyriai prieš ir po treniruotės ar žaidimo suteikia komandą ar žaidėjus remiantiems prekių ženklams dar daugiau dėmesio. Kalbama ne tik apie žaidėjų uniformas, bet taip pat ir apie prekių ženklų afišas dedamas ant sienos ar ekrano už žaidėjų, kai jie yra kalbinami. Taigi, logotipai yra visur prieš, per ir po daugumos prestižinių sporto renginių.

Atsižvelgiant į visa tai, akivaizdu, kad dideli kiekiai kompanijų remia sporto renginius ir nori būti matomi. Tačiau nesunku suprasti, kad skirtingų kompanijų rėmimas atneša ne vienodą naudą. Todėl prekės ženklai ir rinkodaros vadovai ieško būdų, kaip išmatuoti savo prekės ženklo matomumo piniginę vertę visuose pirminiuose kanaluose – transliacijos, srautinio perdavimo ir socialiniuose – naudojant nuoseklią ir patikimą metodiką. Čia atsiranda rėmėjo medijos vertė SMV. SMV yra piniginė prekės ženklo parodymo, remiant socialinį ar transliuojamą turinį, vertė. Pavyzdžiui, ji matuoja, kiek vertingi yra aikštės ženklai NBA žaidimo metu arba MLB komandos paskelbtame socialiniame tinkle. SMV padeda prekių ženklams ir rinkodaros vadovams įvertinti savo investicijas į rėmimą, reklamą ir prekės ženklo partnerystę. **SMV apskaičiuojamas pagal prekės ženklo pasiekiamumą, dažnumą ir kokybę.** Prekės ženklas turi išmatuoti savo SMV, kad suprastų prekės ženklo vertę ir skirtingų rinkodaros kanalų generuojamą ROI [9]. Tam tikri prekių ženklai turi savo automatizuotus vertės skaičiavimo įrankius, gebančius įvertinti jų prekės ženklo matomumą, tačiau dauguma prekių ženklų ar įmonių, teikiančių reklamos paslaugas (taip pat ir Lietuvoje) tai daro rankiniu būdu.



Skaičiuojant visą, kiekvieno sporto rėmėjo žiniasklaidos vertę, atsižvelgiama į keletą veiksnių:

- FMV – tos pačios reklamos vietos ar paskirties vietos, kurioje rodomas prekės ženklas, įsigijimo kaina be jokių rėmimo privilegijų. Jis apskaičiuojamas naudojant vidutines rinkos normas.
- MVP, dar vadinama EMV – atsižvelgia į ekspozicijos kokybę. Tai yra FMV procentinė dalis, priskirta rėmimo poveikiui. Jis skiriasi priklausomai nuo ekspozicijos kokybės, dydžio ir iškilumo.
- trukmės koeficientas (tik vaizdo įrašui) – trukmės koeficientas naudojamas tik vaizdo įrašams ir nustato vertę pagal laiką, kiek prekės ženklas arba jo logotipas buvo matomas ekrane.

SMV apskaičiuojamas padauginus vaizdo įrašų FMV, MVP ir trukmės koeficientą.

Šioje formulėje objektyvus tik vienas parametras – trukmės koeficientas. Šį koeficientą šiuo metu apskaičiuoja skirtingai – vieni naudoja generalizuotas reikšmes, kiti naudoja rankinį skaičiavimo principą. Tam tikri dideli prekės ženklai turi savo įrankius, kurių pagalba seka savo trukmės koeficientą, tačiau didžiajai daliai prekių ženklų tokie įrankiai yra per brangūs.

## 1.2. Atviri logotipų duomenų rinkiniai

Šioje dalyje aptariami patys žymiausi viešai prieinami logotipų duomenų rinkiniai, kurie plačiai naudojami logotipų aptikimo ir atpažinimo modeliams mokytį. Bendras jų visų aprašymas pateiktas lentelėje (žr. 1 lentelė.)

1 lentelė. Atvirų duomenų rinkinių aprašymas

Duomenų rinkinys	Logotipai	Prekių ženklai	Paveikslėliai	Objektai	Metai
FlickrLogos-27 [10]	27	27	1 080	4 671	2011
FlickrLogos-32 [11]	32	32	2 240	5 644	2011
BelaLogos [12]	37	37	10 000	2 695	2009
SportLogo [13]	31	31	2 836	2 836	2020
QMUL-OpenLogo [14]	352	352	27 083	-	2018
FoodLogoDet-1500 [15]	1 500	-	99 768	145 400	2021
LogoDet-3K [16]	3 000	2 864	158 652	194 261	2020

### 1.2.1. FlickrLogos su 27 klasėmis

Duomenų rinkinį sudaro 27 klasės, kiekviena klasė atitinka vieną prekės ženklą. Prekių ženklų nuotraukos atitinka skirtingus kriterijus: paveikslėliai atitinka įvairias apšvietimo sąlygas (natūralioje šviesoje, apšviesti), yra įvairių temų, atvaizduoti iš skirtingų proporcijų, skirtingo dydžio. Duomenų rinkinį iš viso sudaro 1080 paveikslėliai.

### 1.2.2. FlickrLogos su 32 klasėmis

Duomenų rinkinį sudaro 32 klasės, kiekviena klasė atitinka vieną prekės ženklą. Prekių ženklų nuotraukos kaip ir FlickrLogos-32 sudarytos remiantis įvairiomis sąlygomis. Duomenų rinkinį iš viso sudaro 8240 paveikslėliai.

### 1.2.3. *BelgaLogos*

Duomenų rinkinį sudaro prekių ženklų nuotraukos, kurių dydis – 800 pikselių x 800 pikselių. Duomenų rinkinį sudaro 37 skirtingi logotipai, iš viso yra 10000 paveikslėlių.

### 1.2.4. *Logo-Net*

*Logo-Net* tai didelio masto logotipų vaizdų duomenų bazė, kurią sudaro realių produktų vaizdai ir yra skirta logotipo aptikimo ir prekės ženklų atpažinimo modeliams apmokyti. *Logo-Net* sudaro du duomenų rinkiniai:

- „logos-18“: 18 logotipų klasių, 10 prekių ženklų ir 16 043 logotipų objektai.
- „logos-160“: 160 logotipų klasių, 100 prekių ženklų ir 130 608 logotipų objektai.

### 1.2.5. *LogoDet-3K*

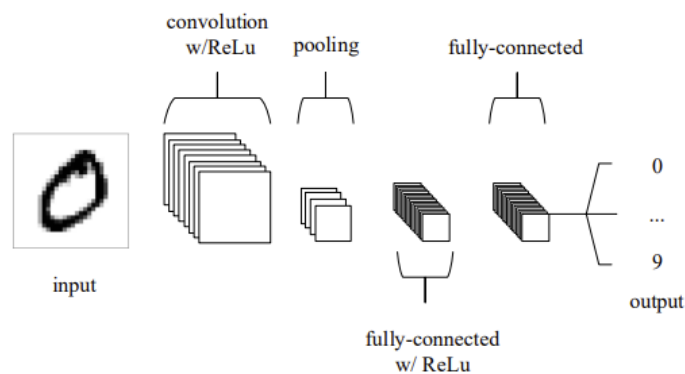
*LogoDet-3K* duomenų rinkinys – vienas iš didžiausių logotipų aptikimo duomenų rinkinių su anotacijomis, kuriame yra 3000 logotipų kategorijų, beveik 200 000 aukštos kokybės rankiniu būdu komentuojamų logotipų objektų ir 158 652 vaizdai. Šis duomenų rinkinys, dėl prekės ženklų gausos, nustato sudėtingesnę etaloną ir gali būti panaudojamas daugeliui esamų lokalizacijai jautrių užduočių, susijusių su logotipų aptikimu, taikymui.

## 1.3. Giliuoju mokymusi grįstos objektų aptikimo ir atpažinimo architektūros

Šioje dalyje aptariamos pačios populiariausios objektų aptikimo ir atpažinimo modelių architektūros. Taip pat išskiriamos šių architektūrų stipriosios savybės bei kur ir kokio tipo uždaviniams jos spręsti yra tinkamiausios.

### 1.3.1. Konvoliuciniai neuroniniai tinklai

Konvoliucinis neuroninis tinklas – vienas iš dirbtinių neuroninių tinklų tipų, kuris labiausiai specializuojasi į objektų aptikimo ir atpažinimo užduotis. Konvoliuciniai neuroniniai tinklai sudaryti iš daugybės sluoksnių, kur kiekvienas sluoksnis geba išmokyti tam tikras paveikslėlio savybes. Jų architektūrą sudaro įvesties sluoksnis, išvesties sluoksnis ir paslėpti sluoksniai. Šie sluoksniai atlieka operacijas, kurios pakeičia duomenis, siekdamas išmokyti duomenims būdingų mokymosi funkcijų. Trys iš labiausiai paplitusių sluoksnių yra konvoliucija, aktyvinimas arba ReLU ir telkimas [17].



1 pav. Bendra CNN architektūra

Nepaisant santykinai mažo sluoksnių skaičiaus, kurio reikia sudaryti CNN, nėra vieno geriausio nustatyto būdo kaip suformuluoti tinkamą. Išanalizavus susijusią literatūrą akivaizdu, kad panašiai kaip ir kitų formų dirbtiniai neuroniniai tinklai, CNN taip pat dažniausiai vadovaujasi bendra architektūra (žr. **1 pav.**), kuri sudaryta iš kelių konvoliucinių pakopų ir vieno ar daugiau visiškai sujungtų sluoksnių, kurie suteikia galutinę išvestį – klasifikavimo modulį. Taigi, pagrindinę tipinę CNN architektūrą sudaro šie komponentai [17]:

Filtrų bankas arba branduoliai: kiekvienas filtras ar branduolys siekia aptikti tam tikrą charakteristiką kiekvienoje įvesties vietoje, todėl erdvinis įvesties vertimas iš būdingo aptikimo sluoksnio bus perkeltas į išvestį be pakeitimų.

Konvoliucijos sluoksnis: konvoliucijos operacija plačiai naudojama skaitmeniniame vaizdų apdorojime, kai 2D matrica, vaizduojanti vaizdą (I), yra sujungta su mažesne 2D branduolio matrica (K) (žr. 1.1).

$$S_{i,j} = (I * K)_{i,j} = \sum_m \sum_n I_{i,j} \cdot K_{i-m,j-n} \quad (1.1)$$

Konvoliucijos procese mažas slankusis filtras eina iš kairės į dešinę bei iš viršaus į apačią per visą vaizdą. Išvesties žemėlapių matmenys gaunami mažesni nei įvesties. Esant reikalui išlaikyti tuos pačius plokštumos matmenis, galima pritaikyti užpildymo techniką, pridėdami nulius aplink įvestį ir pritaikydami branduolio centrą ant atokiausių elementų. Be to, žingsnis reiškia praėjimo tarp dviejų nuoseklių branduolio pozicijų dydį. Dažniausiai žingsnis pasirenkamas lygus 1.

Netiesinio aktyvinimo funkcija: filtrų bankui sugeneravus išvestį, prie jos pritaikoma netiesinė aktyvinimo funkcija, kurios pagalba sudaromi aktyvinimo žemėlapiai. Ši funkcija nustato neurono išvesties elgesį. Yra keletas skirtingų, populiarių aktyvacijos funkcijų. Keletas jų:

*ReLU* (angl. *Rectified Linear Unit function*) – viena iš populiariausių aktyvacijos funkcijų konvoliuciniuose neuroniniuose tinkluose. Tai funkcija, nepaliekanti neigiamų verčių (žr. 1.2).

$$f(x) = \max(0, x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (1.2)$$

Sigmoid funkcija – S formos kreivė. Funkcija svyruoja tarp [0, 1], todėl plačiai naudojama nustatyti tikimybei iš įvesties (žr. 1.3).

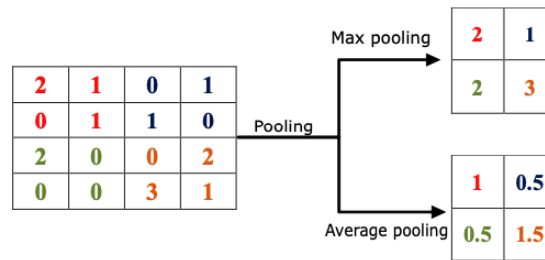
$$f(x) = \frac{1}{1 + e^{-x}} \quad (1.3)$$

Hiperbolinis tangentas – funkcija, panaši į sigmoid, tik intervalas tarp [-1, 1] (žr. 1.4).

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (1.4)$$

Telkimo sluoksnis (angl. *Pooling*): sumažina tinklo parametrų skaičių, sumažindamas konvoliucinių išėjimų erdvinį dydį (žr. **2 pav.**) Be to, telkimo operacijos padeda gauti kintamą mažų įvesties vertimų

atvaizdavimą. Maksimalus telkimas apskaičiuoja maksimalią kiekvieno įvesties pataiso reikšmę, vidutinis telkimas – vidurkius.



2 pav. Telkimo sluoksnio veikimo vizualizacija

Iškritimo sluoksnis: tai reguliavimo sluoksnis, kuris atsitiktinai išmeta tinklo neuronų vienetus, neleidžiant jiems per daug prisitaikyti. Atmetimo technika leidžia susidurti su persimokymo problema, tuo pačiu taip pat pagerina ir tinklo našumą. Jis gali būti taikomas bet kuriam tinklo sluoksniui.

Visiškai prijungtas (FC, angl. *Fully connected*) sluoksnis: konvoliucinių etapų išvestis išlyginama į 1D masyvą ir prijungta prie visiškai sujungto sluoksnio. FC sluoksniai paima konvoliucijos / sujungimo proceso rezultatus ir naudoja juos vaizdui klasifikavimui.

### 1.3.1.1. Konvoliucinių neuroninių tinklų mokymo procesas

Mokymo proceso tikslas – optimizuoti neuroninio tinklo sluoksnių parametrus taip, kad būtų sumažintas skirtumas tarp nurodytų etikečių mokymo duomenų rinkinyje bei išvesties prognozėse. Paprastai atgalinio skleidimo algoritmas yra labiausiai paplitęs neuroninių tinklų mokymo metodas. Mokymo procesas su dauginimu atgal yra toks:

1. pasirinkamas mokymui skirtas vaizdų duomenų rinkinys;
2. perduodama kiekviena partija per tinklą ir gaunama išvestis;
3. apskaičiuojama paklaida tarp nurodytų etikečių ir išvesties prognozių panaudojant praradimo funkciją;
4. paskleidžiama klaida visame tinkle taikant atgalinio skleidimo algoritmą;
5. atnaujinami svoriai, kad būtų sumažinama klaida;
6. kartojama, kol pradės konverguoti arba bus pasiekta iteracijų riba.

Labai svarbu pasirinkti tinkamą duomenų rinkinį. Visų pirma, atkreipiamas dėmesys į paveikslėlių skaičių. Jeigu duomenų rinkinyje jų yra nepakankamai, modelis negebės apsimokyti ir pateikti teisingų rezultatų. Taip pat duomenų rinkinys turi būti subalansuotas – kiekviena klasė turi turėti panašų kiekį paveikslėlių. Duomenų rinkinys padalijamas į tris dalis: mokymo, validacijos bei testavimo. Mokymo dalis naudojama modelio mokymo metu, validacijos – modelio vertinimui mokymo metu bei testavimo – modelio vertinimui po mokymo. Dauguma CNN sistemų reikalauja, kad visi mokymo duomenys būtų vienodos formos (t.y. tų pačių matmenų). Todėl išankstinis duomenų apdorojimas turėtų būti pirmas žingsnis prieš mokymo procesą siekiant normalizuoti duomenis.

### 1.3.1.2. Konvoliucinių neuroninių tinklų trūkumai

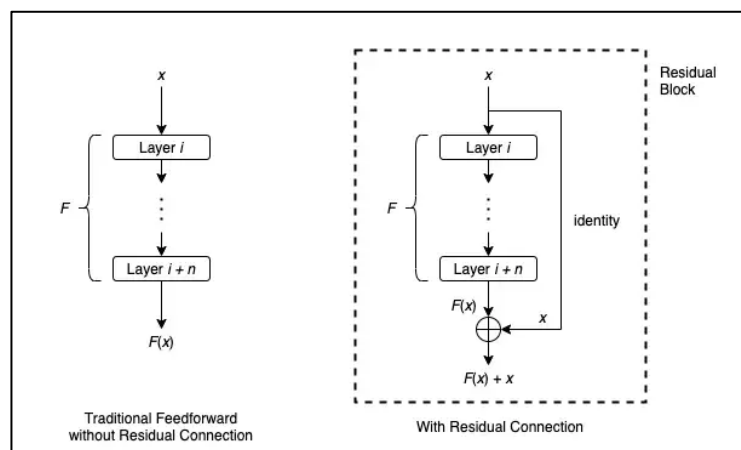
Norėdami gauti tikslesnius objektų aptikimo ar atpažinimo rezultatus, sprendimų, paremtų standartine CNN architektūra, nepakanka. Viena iš pagrindinių tradicinių CNN problemų yra nykstančių gradientų

problema. Tai atsitinka treniruojant labai gilius neuroninius tinklus, kur gradientai, naudojami svoriams atnaujinti treniruotės metu, tekėdami per sluoksnius tampa vis mažesni. Dėl to tinklui sunku mokytis ir gali prastai veikti. Taip pat standartinė CNN architektūra, dėl didelio kiekio mokymui naudojamų resursų, ne visais atvejais yra palanki, norint apmokyti modelį, nes susiduriama su našumo problemomis. Remiantis tuo, dažniausiai tikslo siekiama renkantis pažangesnes architektūras. Kadangi prekių ženklų atpažinimo problemą sprendžiame turėdami didelį pradinių duomenų kiekį, mus domina efektyvūs sprendimai, kurie gebėtų susidoroti su duomenų kiekiu. Taip pat svarbi ir modelio greitaveika, todėl architektūra turi būti ir naši. Šiomis savybėmis standartinė CNN architektūra nepasižymi, todėl šiam uždaviniui spręsti reikalingas modernesnis jos įgyvendinimas.

### 1.3.2. ResNet – liekamasis neuroninis tinklas

ResNet yra konvoliucinio neuroninio tinklo tipas, kuris gali apmokyti labai gilius neuroninius tinklus, sumažindamas nykstančių gradientų problemą. Tradiciniame konvoliuciniame neuroniniame tinkle funkcijos išgaunamos įvestis perduodant per kelis filtrų sluoksnius – kuomet tinklas gilesnis, tuo sudėtingesnis tampa modelio funkcijų apmokymas. Tačiau tinklui gilėjant, gradientai, naudojami svoriams atnaujinti treniruotės metu, tampa vis mažesni. ResNet modelis sprendžia šią problemą įveddamas liekamuosius ryšius tarp sluoksnių, kurie palengvina gradientų tėkmę tinkle ir leidžia apmokyti daug gilesnius tinklus [18].

Liekamasis ryšys suteikia papildomą kelią duomenims pasiekti kitas neuroninio tinklo dalis praleidžiant kai kuriuos sluoksnius (žr. 3 pav.). Tarkim turint sluoksnių seką nuo  $i$  iki  $i + n$ , kur  $F$  atstoja šiuos sluoksnius atitinkančias funkcijas, tradicinėje tiekimo į priekį architektūroje,  $x$  pereis per visus sluoksnius, tačiau liekamieji ryšiai suteikia galimybę tam tikrus sluoksnius praleisti [19].



3 pav. FFNN ir ResNet palyginimas

#### 1.3.2.1. Skirtingų sluoksnių ResNet

Šiuo metu populiariausios kelios skirtingos ResNet architektūros, kurios tarpusavyje skiriasi savo turimų sluoksnių skaičiumi [18] (žr. 2 lentelė).

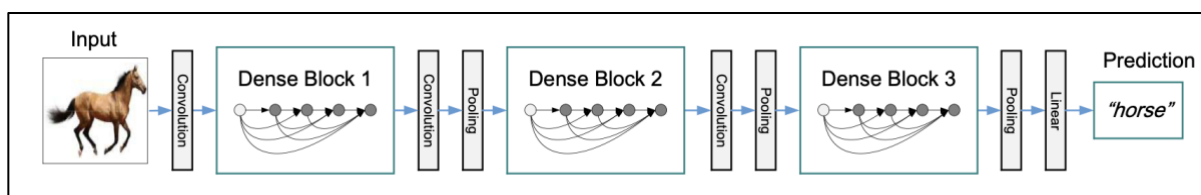
2 lentelė. Skirtingų ResNet palyginimas

Pavadinimas	Aprašymas
ResNet18 ResNet34	Mažesnės ResNet versijos, turinčios mažiau sluoksnių. Paprastai naudojamos užduotims, kurių skaičiavimo ištekliai yra riboti.

ResNet50	Didesnės <i>ResNet</i> versijos, turinčios daugiau sluoksnių. Paprastai jie naudojami sudėtingesnėms užduotims, tokioms kaip vaizdų klasifikavimas ir objektų aptikimas.
ResNet101	
ResNet152	

### 1.3.3. *DenseNet*

*DenseNet* yra konvoliucinio neuroninio tinklo (CNN) architektūros tipas, tiesiogiai vienas su kitu jungiantis visus tinklo sluoksnius. Tai sukuria tankų bloką, kuris leidžia pakartotinai naudoti funkcijas ir padeda sušvelninti nykstančių gradientų problemą. Siekiant sumažinti funkcijų žemėlapių dydį, *DenseNet* naudoja pereinamuosius sluoksnius, sudarytus iš paketinio normalizavimo sluoksnio, po kurio seka  $1 \times 1$  konvoliucija ir  $2 \times 2$  vidutinis telkimas (žr. **4 pav.**). Pereinamasis sluoksnis sumažina aukščio ir pločio matmenis, bet palieka tokius pačius elementų matmenis [20].



4 pav. *DenseNet* tinklo architektūra

Vienas iš šio modelio privalumų yra tai, kad jo mokymui kur kas mažiau parametrų nei tradiciniai konvoliuciniai neuroniniai tinklai, nes nereikia iš naujo mokytis perteklinių funkcijų žemėlapių. Tradicinės, į priekį nukreiptos neuroninių tinklų architektūros (FFNN), būsenas perduoda iš sluoksnio į sluoksnį – kiekvienas sluoksnis nuskaito būseną iš ankstesnio sluoksnio ir rašo į kitą sluoksnį. [21] Architektūroje keičiantis būsenoms, kartu perduodama ir informacija, kurią reikia išsaugoti. *DenseNet* šią informacijos išsaugojimą daro per papildomus tapatybės transformavimus. Kadangi *DenseNet* architektūroje sluoksniai tarpusavy nebendradarbiauja, mokymo metu galima dalį jų atsitiktine tvarka pašalinti. Dėl to *DenseNet* būseną panaši į pasikartojančius neuroninius tinklus (angl. *recurrent neural network*), tačiau parametrų skaičius yra daug didesnis, nes kiekvienas sluoksnis turi savo svorį.

### 1.3.4. *VGGNet*

*VGGNet* yra konvoliucinio neuroninio tinklo architektūra, kurią publikavo Oksfordo universiteto Visual Geometry Group mokslininkai. Architektūrai būdingi nedideli konvoliuciniai filtrai ( $3 \times 3$ ) ir labai gilūs tinklai, kurie gali turėti iki 19 sluoksnių. *VGGNet* architektūra suskirstyta į penkis pagrindinius blokus, kur kiekviename yra keli konvoliuciniai sluoksniai ir vienas ar daugiau maksimalaus telkimo sluoksnių. Pristačius originalią *VGGNet* architektūrą ir atlikus testavimus su „*ImageNet*“ duomenų rinkiniu tuo metu buvo pasiektas vienas iš geriausių vaizdų klasifikavimo našumų. *VGGNet* architektūra išsiskiria tuo, kad naudoja labai mažus konvoliucinius filtrus, kurie leidžia išmokti smulkias savybes [22].

*VGGNet* gali būti išmokytas atlikti specifinius objektus, tačiau dažniau šis modelis naudojamas jau iš anksto paruoštas. Modelis yra išmokytas naudojant „*ImageNet*“ duomenų imtį, todėl jo svoriai gali būti naudojami kaip atspirties taškas treniruojant mažesnę duomenų rinkinį, skirtą konkrečiai užduočiai atlikti, naudojant perkeliamojo mokymosi techniką [23] (perkeliamasis mokymasis yra mašininio mokymosi metodas, kai modelis, parengtas atlikti vieną užduotį, yra pakartotinai naudojamas kaip atspirties taškas modeliui atliekant antrą užduotį).

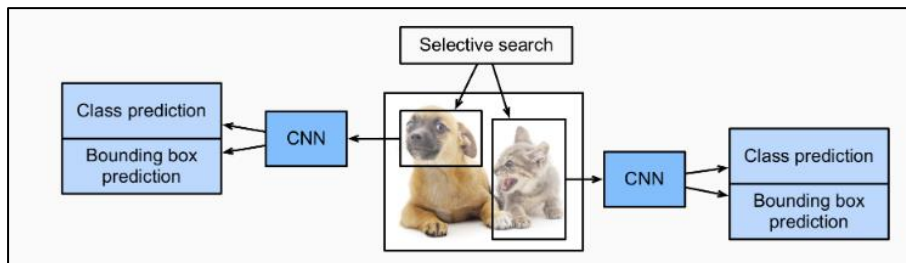
### 1.3.5. Regionais grįsti konvoliuciniai neuroniniai tinklai

Regionais grįstas konvoliucinis neuroninis tinklas yra objektų aptikimo algoritmo tipas, kuris, naudodamas CNN, geba vaizde nustatyti ir rasti išmokus objektus. Paprastai šie algoritmai naudoja slankiojo lango vaizdo nuskaitymo techniką, kurios metu iš paveikslėlio yra išgaunami regionai, kurie vėliau perleidžiami per CNN tam, kad būtų lokalizuoti ir klasifikuoti objektai. Regionais grįsti CNN dažnai laikomi tikslesniais nei kitų tipų objektų aptikimo algoritmai, pvz., vieno šūvio detektoriai. Vienas iš pagrindinių regioninių CNN privalumų yra tai, kad juos galima išmokyti konkrečiai pasirinktai problemai spręsti konkrečioje aplinkoje. Tokie modeliai veikia dideliu tikslumu ir pasižymi pakankamai sparčia greita veika. Dėl savo pritaikomumo šio tipo modeliai plačiai naudojami daugelyje realių laiku veikiančių programų, pvz., savarankiškai važiuojančiuose automobiliuose, saugumo stebėjimo sistemose ir medicininio vaizdų apdorojime [24].

#### 1.3.5.1. R-CNN, greitas R-CNN ir greitesnis R-CNN

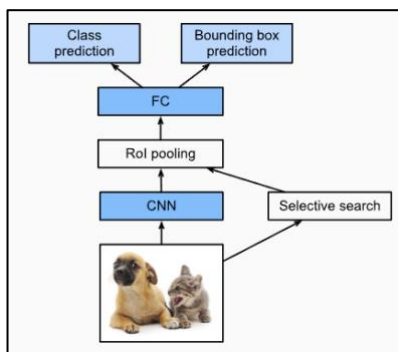
R-CNN, greitas R-CNN ir greitesnis R-CNN yra „Microsoft Research“ sukurti regionais grįsti objektų aptikimo algoritmai. Pradinio R-CNN modelio tikslas buvo priimti vaizdo įvestį ir sukurti ribinių laukų rinkinio išvestį, kur kiekviename laukelyje yra objektas ir objekto kategorija. Tačiau laikui bėgant algoritmas buvo vis tobulinamas, todėl dabar moderniausios algoritmo variacijos gali aptikti objektus net ir trimatėje aplinkoje [25].

R-CNN (Regionai su CNN funkcijomis) galimų objektų regionams aptikti naudoja atrankinės paieškos algoritmą. Kiekvienas aptiktas objekto regionas paleidžiamas per CNN ir yra priskiriamas arba objektui, arba fonui. Nors modelis naudoja iš anksto išmokytą CNN vaizdo savybėms gauti, vis dėlto modelis veikia pakankamai lėtai (žr. **5 pav.**) [25].



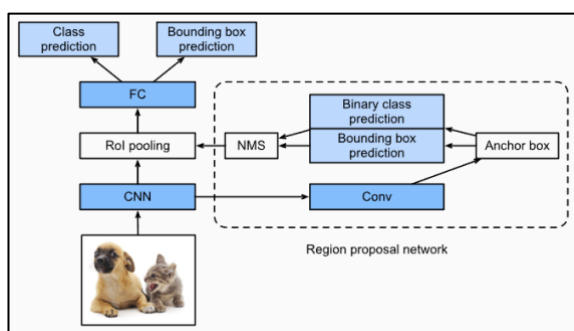
**5 pav.** R-CNN veikimo principo diagrama

Greitas R-CNN pagerina R-CNN dalindamasis skaičiavimais tarp regiono pasiūlymo ir klasifikavimo etapų, todėl bendras procesas tampa efektyvesnis. Užuo paleidęs CNN kiekviename siūlomame regione, visas vaizdas perduodamas per CNN, kad būtų išskirtos funkcijos, kurios vėliau naudojamos klasifikuojant kiekvieną siūlomą regioną (žr. **6 pav.**). [26]



6 pav. Greito R-CNN veikimo principo diagrama

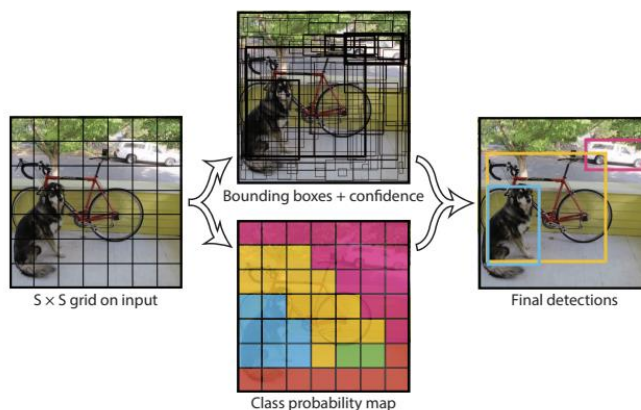
Greitesnis R-CNN dar labiau pagerina greitąjį R-CNN, naudodamas regiono pasiūlymo tinklą (RPN), kad sukurtų siūlomus regionus, o ne naudojant atskirą algoritmą, pvz., atrankinę paiešką [27]. Tai leidžia dar greičiau aptikti objektą (žr. 7 pav.).



7 pav. Greitesnio R-CNN veikimo principo diagrama

### 1.3.6. YOLO

YOLO yra objektų aptikimo realiuoju laiku algoritmas, kurį sukūrė Josephas Redmonas, Santosh Divvala, Ross Girshick ir Ali Farhadi. YOLO padalija vaizdą į langelių tinklą ir kiekvienam tinklelio elementui priskiria besiribojančių langelių rinkinį ir tikimybę, kad langelyje yra objektas. Besiribojančių langelių ir klasių tikimybėms apskaičiuoti naudojamas CNN, kuris vienu perėjimu tiesiai iš viso vaizdo apskaičiuoja šias reikšmes. Šis veikimo principas užtikrina ypač sparčią greitaveiką, todėl algoritmas ypač tinka realaus laiko problemoms spręsti [28]. (žr. 8 pav.).



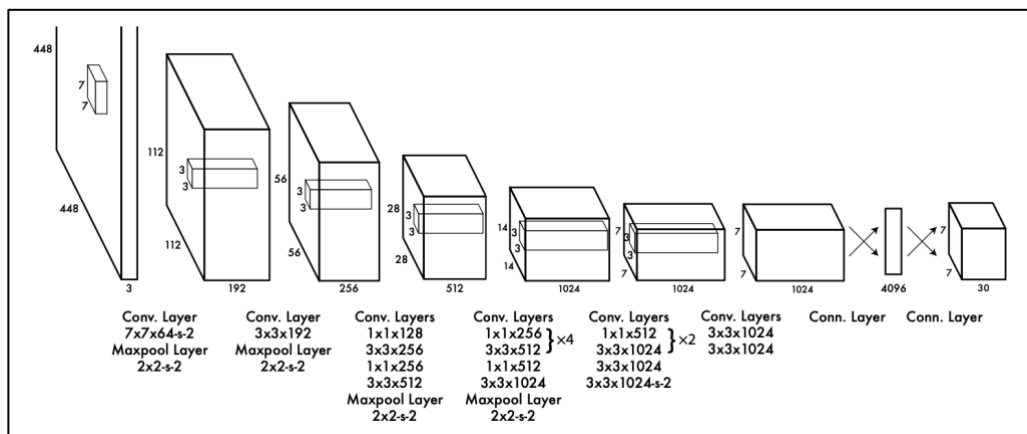
8 pav. YOLO iliustruotas veikimo principas [27]



Skirtingai nei regionais grįstuose metoduose, kuriuose naudojama slankiojančio lango technika objektų regionams aptikti, YOLO mokymo ir testavimo metu mato visą vaizdą, todėl netiesiogiai koduoja kontekstinę informaciją apie klases. Greitas R-CNN neretai klaidingai priskiria foną kaip objektą, nes nemato platesnio konteksto, tik tam tikrą jo dalį. Lyginant su greitu R-CNN, YOLO šioje srityje padaro mažiau nei pusę fono klaidų.

### 1.3.6.1. YOLO architektūra

YOLO architektūrą įkvėpė *GoogLeNet* [29] vaizdų klasifikavimo modelis. Tinklą sudaro 24 konvoliuciniai sluoksniai, po kurių seka 2 visiškai sujungti sluoksniai. Vietoj pradinių modulių, naudojamų *GoogLeNet*, naudojamas  $1 \times 1$  redukcinius sluoksnius, po kurio seka  $3 \times 3$  konvoliuciniai sluoksniai (žr. 9 pav.). Pradiniai konvoliuciniai tinklo sluoksniai išskiria iš vaizdo ypatybes, o visiškai sujungti sluoksniai numato išvesties tikimybes ir koordinatas.



9 pav. YOLO architektūra

YOLO objektų rėžių prognozėms nustato stiprius erdvinis apribojimus, nes kiekvienas tinklo langelis gali turėti tik vieną klasę. Šis erdvinis apribojimas riboja netoliese esančių objektų aptikimo skaičių. Taip pat modelis kur kas sunkiau aptinka smulkius objektus, ypač jeigu jie persidengia. Svarbu paminėti, kad objektų rėžių numatymui modelis negali interpretuoti skirtingų objektų konfigūracijų ar formato, modelis geba numatyti objektus tik pagal tai, kokie jie yra mokymo duomenų imtyje [28].

Skirtingos YOLO versijos naudoja skirtingus modelius paveikslėlio savybėms išgauti. Pavyzdžiui, YOLOv3 architektūrą papildė *DarkNet-53* konvoliucinis neuroninis tinklas, kurį sudaro net 53 sluoksniai. Lyginant su ankstesnės versijos YOLOv2 naudojamu *DarkNet-19*, *DarkNet52* yra kur kas tikslesnis, nes jį sudaro beveik 3 kartus daugiau sluoksnių. Tačiau dėl modelio dydžio nukenčia greitaveika [30].

## 1.4. Logotipų aptikimo ir atpažinimo sprendimų analizė

Šioje dalyje aptariamos architektūros, kuriomis remiantis literatūros autoriai aprašė sukurtus prekių ženklų aptikimo ar atpažinimo modelius. Taip pat aptariami gauti rezultatai bei modelių privalumai ar trūkumai.

### 1.4.1. Logotipų atpažinimas naudojant *ResNet*

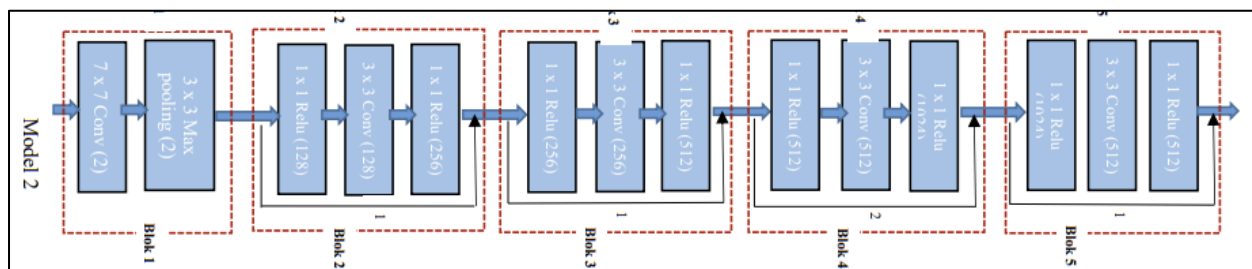
Logotipų aptikimo problemai spręsti autorius naudoja *ImageNet* duomenų imtį bei pritaiko skirtingas *ResNet* versijas ir pastebi, kad tam tikros *ResNet* variacijos (*ResNet18*, *ResNet34*) nėra pritaikytos

išsikeltam uždaviniui spręsti ir pateikia tikslumą, žemesnį nei 0,5. Kitose testuotose variacijose (*ResNet50*, *ResNet101*, *ResNet152*) susiduriama su kita problema – dėl didelio parametrų skaičiaus, modelio mokymui reikia labai daug skaičiavimo resursų [31].

Pastebėtoms problemoms spręsti pristatoma nauja vieno etapo logotipo detektoriaus sistema, kurioje detektoriaus pagrindas yra paprastas sutelktas tinklas su skirtingomis blokų išdėstymo architektūromis (žr. **10 pav.**). Darbe pateikti rezultatai parodė, kad sukurtas modelis geba aptikti logotipus panašiu tikslumu kaip ir *ResNet50*, tačiau pasižymi spartesne greitaveika, nes mokymui ir aptikimui reikia mažiau skaičiavimo resursų (žr. **3 lentelė**) [30].

**3 lentelė.** *ResNet50* ir autoriaus pasiūlyto modelio palyginimas

	<i>ResNet50</i>	Sukurtas modelis
Parametrų skaičius	36,797,857	22,392,225
mAP	0.382 ± 0.0921	0.408 ± 0.1050
Testavimo laikas	41.18 s.	36.45 s.

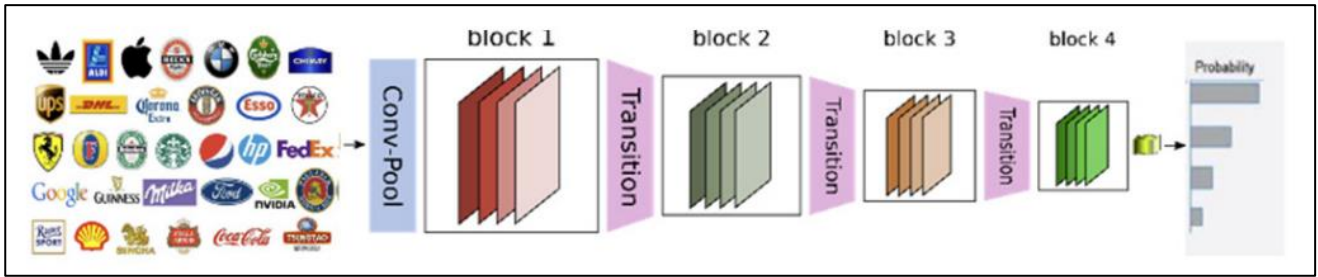


**10 pav.** Autoriaus pasiūlyto modelio architektūra

#### 1.4.2. Logotipų atpažinimas naudojant *DenseNet*

Darbe, kuriame logotipų aptikimas įgyvendintas naudojant *DenseNet*, siekiant įvertinti siūlomą modelį, mokymui ir testavimui buvo naudojamas *FlickrLogos-32* logotipo atpažinimo duomenų rinkinys. *FlickrLogos-32* yra viešai prieinamas duomenų rinkinys, kuriame yra 8240 realaus pasaulio 32 skirtingų prekės ženklų logotipų rinkinys. Jis buvo sukurtas logotipų paieškos ir kelių klasių logotipų aptikimo ir atpažinimo sistemoms įvertinti. Duomenų rinkinys buvo suskirstytas į tris atskirus pogrupius: P1, P2 ir P3. Kiekviename pogrupyje yra visų 32 klasių vaizdai. Mokymo rinkinį P1 sudaro 320 vaizdų, 10 vaizdų vienoje klasėje, o validavimo rinkinį P2 ir testų rinkinį P3 sudaro 3960 vaizdų. *DenseNet* modelio rezultatai buvo 4% geresni, nei kitas aukščiausių rezultatą pateikęs modelis [32].

Sprendimui siūlomą modelį sudaro keturi blokai: prieš pirmąjį tankų bloką  $7 \times 7$  konvoliucijos ir  $3 \times 3$  vidutinio telkimo sluoksniai, kurie skirti iš įvesčių išskirti funkcijas. Tarp kiekvieno tankaus bloko - pereinamasis sluoksnis, sudarytas iš  $1 \times 1$  konvoliucijos ir  $2 \times 2$  vidutinio telkimo. Ties ketvirtu bloku seka  $7 \times 7$  vidutinis telkimo sluoksnis, skirtas nustatyti objektų žemėlapių dydį kuris jungsis su pilnai sujungtu sluoksniu (žr. **11 pav.**) [32].



11 pav. DenseNet interpretacija logotipų atpažinimo uždaviniui spręsti

### 1.4.3. Greito R-CNN ir greitesnio R-CNN taikymas prekių ženklų aptikime

Prekių ženklų aptikimui darbe pasirinkti greitas R-CNN ir greitesnis R-CNN algoritmai. Eksperimentai atliekami su viešai prieinamu *FlickrLogos-32* duomenų rinkiniu, kurį sudaro 32 skirtingų prekių ženklų logotipai. Skirtingai nuo logotipo atpažinimo, logotipui aptikti naudojami tik priekinio plano vaizdai, kuriuose yra tam tikro tipo logotipo objektas.

Greitas R-CNN ir greitesnis R-CNN tinklai yra iš anksto paruošti naudojant *ImageNet* duomenų rinkinį. Kiekvieno vaizdo dydis keičiamas taip, kad jo trumpiausia kraštinė būtų 600, o ilgiausia kraštinė neviršytų 1000 pikselių. Mokymo metu, regionų pasiūlymai, kurių sankirta virš sąjungos (IoU) sutampa su duomenų imtyje nurodytais objekto rėžiais (angl. *bounding box*) daugiau nei 0,5, yra laikomi aptiktais objektais, o kiti – fonu. Modelio vertinimui naudojama mAP reikšmė.

Darbe pateiktuose rezultatuose galime pastebėti, kad tikslumas varijuoja ties skirtingais prekių ženklais. Tam tikrus prekių ženklus modeliai geba aptikti labai tiksliai (tikslumas siekia 90%), tačiau kai kurių prekių ženklų aptikimas tesiekia iki 60% [33].

Analizuojant neteisingai aptiktus logotipų atvejus, nustatyta kad dažniausiai pasitaiko dviejų tipų logotipai, kuriuos sunku aptikti:

- paprasti logotipai – jei logotipai yra pernelyg minimalistiniai ir tinklui nepateikiama pakankamai informacijos, kad galėtų juos atskirti;
- logotipai, kurie turi daug skirtingų variacijų – kai kurie logotipai gali akivaizdžiai pasikeisti dėl skirtingo dizaino stiliaus arba esančių skirtingose prekėse, o tai kelia didelių iššūkių patikimam logotipo aptikimui.

### 1.4.4. Logotipų aptikimas naudojant YOLO

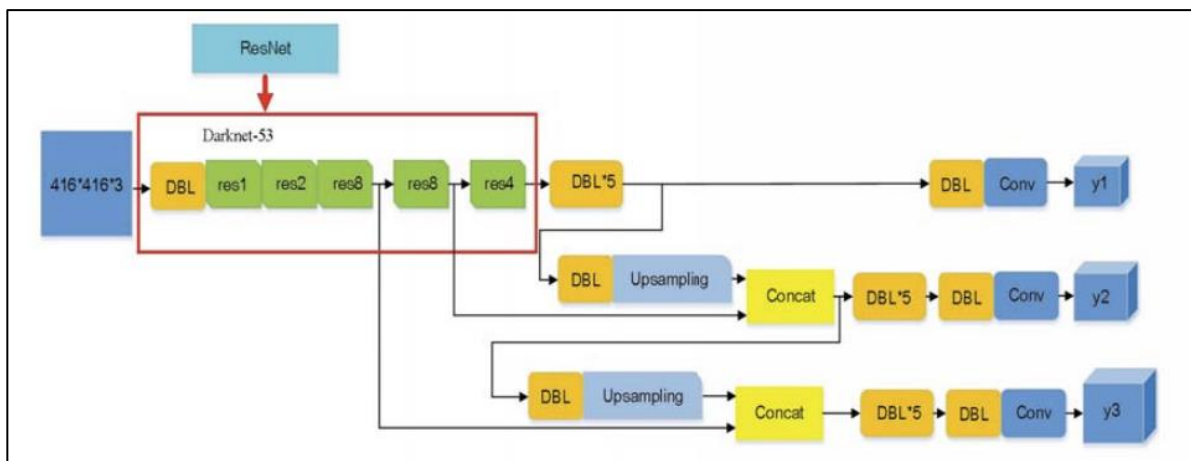
Logotipų aptikimo analizei pasitelktas YOLOv2. Jį sudaro 19 konvoliucinių sluoksnių bei 5 maksimalaus telkimo sluoksniai. Darbe mokymui ir testavimui naudota duomenų imtis, kurią autorius susikūrė pats atlikęs didelį kiekį pasirinktų prekių ženklų nuotraukų bei naudodamas internetinius šaltinius. Renkant šiuos vaizdus, atsižvelgta į skirtingas apšvietimo sąlygas, skirtingus kampus, skirtingą fotoaparato aiškumą ir skirtingą foną. Kiekvienai kategorijai iš viso buvo surinkta 1500 vaizdų, iš kurių 1000 vaizdų surinkti iš skirtingų parduotuvių, o likusieji 500 paimti iš interneto išteklių. Tikslumui įvertinti pasitelktos IoU vertės. Darbe pasiektas 96% bendras tikslumas, o kai kuriuos prekių ženklus modelis geba aptikti net 99% tikslumu [34].

## 1.5. Objektų aptikimo ir atpažinimo sprendimų analizė

Skirtingi giliojo mokymosi modeliai, gebantys atlikti objektų aptikimą ar atpažinimą, turi stipriąsias ir silpnąsias savybes. Siekiant aukščiausio tikslumo rezultatų, neretai pasitelkiami ir hibridiniai giliojo mokymosi modeliai, kurie apjungia atskiras skirtingų modelių dalis ir pabrėžia kiekvieno stipriąsias savybes tam tikrai užduočiai spręsti. Šiai užduočiai aktualiausi sprendimai, kurie geba kuo tiksliau aptikti įvairius objektus (idealiausiu atveju – logotipus). Toliau aptarsime literatūros autorių aprašytas objektų aptikimo ir atpažinimo architektūras, kurias autoriai įgyvendino remdamiesi skirtingais giliojo mokymosi modeliais.

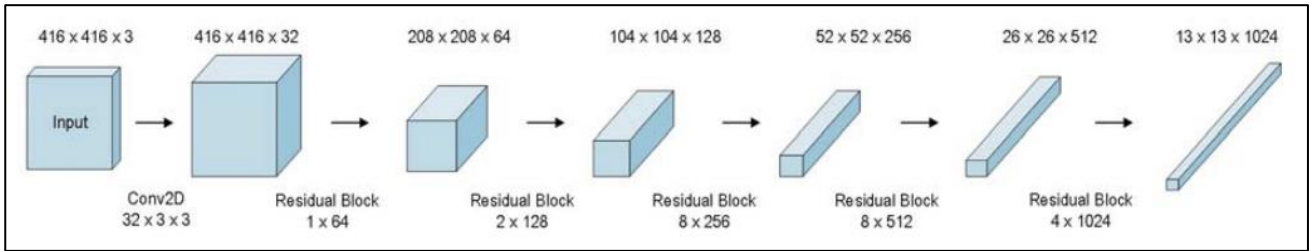
### 1.5.1. *ResNet* ir YOLO hibridinė sistema

Nors YOLOv3 algoritmas buvo patobulintas ir geba aptikti skirtingų mastelių objektus, tai ne visai išsprendė labai smulkių objektų aptikimo problemas, tikslumas vis dar išlieka gana prastas, o atpažinimo klaida didelė. Darbo autorius nusprendė, kad geriausias būdas dar labiau pagerinti modelio tikslumą yra suprojektuoti dar gilesnės architektūros modelį, tačiau tai didina gradiento išnykimo problemas atsiradimą. Šios problemos sprendimui pasitelktas *ResNet* modelis: remiantis *Darknet-53* tinklo struktūra vaizdo savybių išgavimui, vietoj jo pritaikomas *ResNet* modelis vaizdo savybėms išgauti [35].



12 pav. *ResNet* ir YOLO hibridinė architektūra

Pirmiausia, įvedami 3 kanalų,  $416 \times 416$  pikselių vaizdai, iš kurio naudojant *Darknet-53* tinklą ir *ResNet* išgaunamos funkcijos. Toliau lokali ypatybių sąveika realizuojama naudojant konvoliucijos branduolį, po kurio atliekamas kelių mastelių prognozės. Paveikslėlyje (žr. 12 pav.) DBL modulis atitinka dvimatį konvoliucinį sluoksnį, paketinio normalizavimo sluoksnį ir *LeakyReLU* sluoksnį, kurie kartu sudaro mažiausią tinklo komponentą. Tokiu būdu aktyvavimo funkcija paskirstoma tiesiniu intervalu, o modelis drąsiau leidžiasi gradientu [35].



13 pav. DarkNet-53 architektūra

Hibridinis YOLO ir *ResNet* tinklas funkcijų išskyrimui naudoja *Darknet-53* tinklą (žr. 13 pav.) kartu su *ResNet* tinklu. Pirma, *Darknet-53* naudoja  $3 \times 3$  ir  $1 \times 1$  konvoliuciją tam, kad sumažintų objektų žemėlapio matmenis bei nustatytų žingsnį. Tačiau šiuo būdu nėra išgaunamos galutinės funkcijos, nes objekto padėties identifikavimo tikslumas yra menkas, o atkūrimo dažnis yra mažas. Todėl įvedamas papildomai *ResNet*, kurio pagalba išgaunamos papildomos funkcijos. Galiausiai apskaičiuojamas abiejų modelių išvesties vidurkis ir taip pasiekiamas bendras funkcijų gavimas.

Modelio testavimui naudota VOC2007 duomenų imtis, kurią sudaro daugybė vaizdų, kurie yra pritaikyti objektų aptikimo ar atpažinimo uždaviniui spręsti. Šiame darbe pasirinkti 5000 paveikslėlių, kuriuose skirtingomis sąlygomis atvaizduoti tam tikrų kategorijų objektai – arkliai, traukiniai, katės ir pan. (20 skirtingų). 250 paveikslėlių buvo naudojami modelio validacijai. Kadangi užduotis sprendžia daugelio skirtingų objektų kategorijų aptikimą, standartinių vertinimo metrikų, kurios naudojamos kai modelis geba aptikti tik vienos žymos objektus, nepavyks. Autorius darbe naudoja mAP vertinimo rodiklį.

$$\text{Tikslumas} = \frac{TP}{TP + FP}$$

$$\text{Vidutinis tikslumas} = \frac{\sum \text{Tikslumas}}{n} \quad (1.5)$$

$$\text{mAP} = \frac{\text{Vidutinis tikslumas}}{n}$$

Modelio rezultatai pateikti lentelėje (žr. 4 lentelė). Iš rezultatų galime matyti, kad hibridinis Yolo – *ResNet* modelis pasižymi geresniu tikslumu, tačiau dėl papildomų skaičiavimų nukenčia greitaveika.

4 lentelė. Skirtingų testuotų modelių rezultatų palyginimas

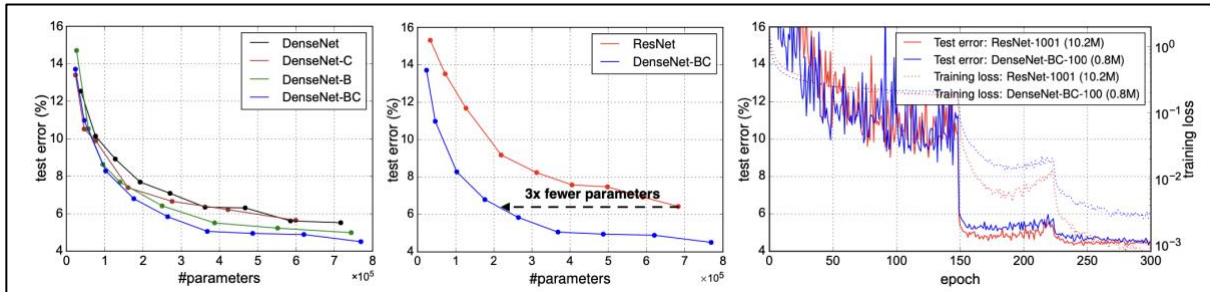
	Greitesnis R-CNN	YOLOv3	SSD	Yolo - <i>ResNet</i>
mAP	69.91	73.02	70.87	75.36
Laikas, s	260	350	220	380

### 1.5.2. Objektų aptikimas naudojant *DenseNet*

*DenseNet* teikiamus privalumus galime pastebėti atliktuose bandymuose, kuriuose gauti rezultatai lyginami su kitų modelių teikimais rezultatais [21]. Pastebėta, kad *DenseNet* yra linkęs nuosekliai pagerinti tikslumą didėjant parametru skaičiui, be jokių veikimo pablogėjimo ar per didelio prisitaikymo požymių. Atliktuose testavimuose, naudojant įvairius konkurencingus duomenų rinkinius

su įvairiais parametrais, modelis pasiekė kur kas geresnius rezultatus nei panašūs modeliai. Be to, šiems rezultatams pasiekti, *DenseNet* reikalauja kur kas mažiau parametru ir mažiau skaičiavimų.

Straipsnio autoriaus pateiktuose rezultatuose galime pastebėti, kad *DenseNet* pateikia labai panašius rezultatus kaip ir *ResNet*, tačiau veikia našiau, nes modelį sudaro žymiai mažiau parametru [21]. Pavyzdžiui, modelis *DenseNet-201* su 20M parametru duoda panašią paklaidą kaip ir *ResNet101* su daugiau nei 60M parametru (žr. **14 pav.**).



**14 pav.** *ResNet* ir *DenseNet* rezultatų palyginimas [20]

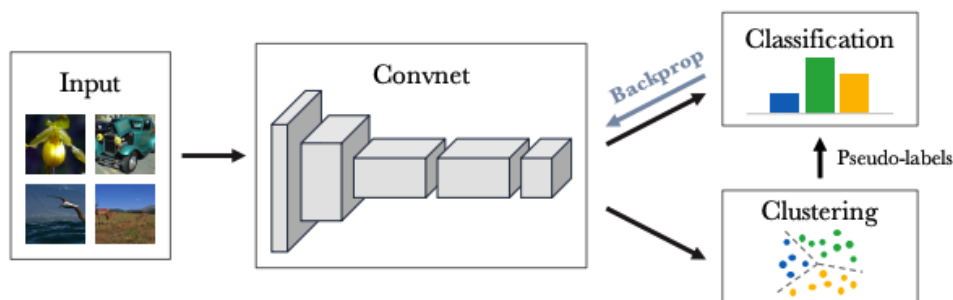
## 1.6. Giliuoju mokymosi grįstas panašių objektų grupavimas

Neprižiūrimi klasterizacijos metodai priklauso mašininio mokymosi metodų kategorijai, skirtai panašioms duomenų taškams sugrupuoti pagal būdingus modelius ar panašumus be išankstinio duomenų žymėjimo. Pagrindinis šių metodų tikslas yra atrasti paslėptas struktūras arba ryšius duomenų rinkiniuose ar suskirstant duomenis į atskiras grupes. Vienas ryškus pavyzdys yra *k-means* klasterizacijos algoritmas, kuris suskirsto duomenis į  $k$  grupių, sumažindamas dispersiją kiekviename klasteryje [36]. Kiti metodai apima hierarchinį klasterizavimą, tankiu pagrįstą grupavimą ir spektrinį grupavimą, kurių kiekvienas turi savo stipriąsias puses ir panaudos atvejus. Neprižiūrimas grupavimas atlieka lemiamą vaidmenį įvairiose srityse, tokiose kaip modelio atpažinimas, duomenų tyrinėjimas ir anomalijų aptikimas, suteikiant vertingų įžvalgų apie sudėtingų duomenų rinkinių pagrindines struktūras, nepasikliaujant iš anksto nustatytais etiketėmis. Be to, neprižiūrimi grupavimo metodai padeda sumažinti matmenis ir išgauti ypatybes, padedant identifikuoti esminius komponentus didelės apimties duomenų rinkiniuose. Šie metodai ypač naudingi tais atvejais, kai tikrosios klasės etiketės nežinomos arba jas sunku gauti. Pavyzdžiui, hierarchiniai klasterizavimo metodai sukuria į medį panašią klasterių struktūrą, atskleidžiančią ryšius ne tik tarp atskirų duomenų taškų, bet ir skirtinguose hierarchijos lygiuose. Tankiu pagrįsti grupavimo metodai, tokie kaip DBSCAN [37], gali nustatyti įvairių formų ir dydžių grupes pagal duomenų tankį. Spektrinis grupavimas panaudoja giminingumo matricų savąsias reikšmes, kad sugrupuotų panašius duomenų taškus, dažnai sukuriant reikšmingas grupes netiesiškai atskiriamuose duomenų rinkiniuose. Neprižiūrimi klasterizacijos metodai ir toliau yra labai svarbūs atskleidžiant įvairių duomenų tipų modelius ir struktūras, todėl jie yra būtini įrankiai platesnėje mašininio mokymosi ir duomenų analizės aplinkoje.

*K-means* algoritmas – vienas iš dažniausiai naudojamų klasterizavimo metodų. Literatūroje siūloma įvairių *k-means* plėtinių. Nors tai yra neprižiūrimo mokymosi algoritmas, *k-means* algoritmas ir jo plėtiniai veikia naudodami iš anksto nustatytais parametrais. Literatūroje siūloma naujų schemų, kurios padeda išvengti išankstinio klasterių skaičiaus pasirinkimo. Viename iš straipsnių aprašoma strategija, kurioje naudojamas *U-k-means* algoritmas pradinio klasterių kiekio pasirinkimui [38]. Iteracijų metu, *U-k-means* algoritmas, naudodamas pateiktą duomenų struktūrą, padeda panaikinti nereikalingus klasterius. *U-k-means* pranašumai yra pradinio nustatymo ir parametru laisvumas, taip pat atsparumas

skirtingiems klasterių kiekiams ir formoms su automatišku klasterių skaičiaus nustatymu. *U-k-means* algoritmas buvo testuojamas su keliais sintetiniais ir realiais duomenų rinkiniais, taip pat rezultatai buvo palyginti su dauguma esamų algoritmų, tokių kaip R-EM, C-FS, *k-means* su nustatytu klasterių skaičiumi  $c$ , *k-means + gap* ir *X-means* algoritmai. Rezultatai parodo *U-k-means* klasterizavimo algoritmo pranašumą.

Autorius darbe pristato *DeepCluster* – klasterizacijos metodą, kuris naudodamas konvoliuninius neuroninius tinklus, geba priskirti duomenis klasteriams pagal gautus parametrus [39] (žr. **15 pav.**). *DeepCluster* iteratyviai sugrupuoja funkcijas su standartiniu klasterizacijos algoritmu, *kmeans*, ir naudoja paskesnius priskyrimus kaip priežiūrą, kad atnaujintų rezultatus. Testuojant metodą su dideliu duomenų rinkiniu, tokiu kaip *ImageNet* ar YFCC100M, jis pasiekia geresnius rezultatus nei ankstesnės geriausiomis laikytos sistemos. Šis metodas daro nedaug prielaidų apie įvestis ir nereikalauja daug srities specifinių žinių, todėl jis yra palankus objektų aptikimo modelio kūrimui.



**15 pav.** *DeepCluster* architektūra [38]

## 1.7. Logotipų aptikimo problemos

Šiame skyriuje aprašomos dažniausiai literatūroje sutiktos logotipų aptikimo problemos, kurios daro didžiausią įtaką gaunamiems tikslumo rezultatams.

### 1.7.1. Mažų logotipų aptikimo problema

Lyginant su dideliais logotipais, mažų logotipų aptikimo tikslumas yra kur kas prastesnis ir tai yra dėl kelių priežasčių. Visų pirma, nedideliuose mažos skiriamosios gebos logotipuose yra mažai vaizdinės informacijos, todėl sunku išskirti išskirtines jų savybes. Maži logotipai turi ribotą vizualaus turinio vietą, dėl to detalės gali prarasti aiškumą, o teksto elementai gali tapti sunkiai įskaitomi. Antra, kadangi mažas logotipas apima nedidelį vaizdo plotą, jį supantį langelį rasti sunkiau nei didelį: prognozės procese klaida, kurią sukelia mažo logotipo prognozuojamo aplinkos rėmelio pikselių nuokrypis, yra kur kas mažesnė nei klaida, kurią gauname didelio logotipo atveju. Galiausiai, duomenų rinkiniuose yra kur kas mažiau mažų logotipų pavyzdžių. Mokymo procese, kai ranka nustatytas rėmelis labai skiriasi nuo mažo logotipo tikrosios ribos, teigiamas mažo logotipo mokymo pavyzdžių skaičius bus daug mažesnis nei didelio logotipo atveju. Dėl šios priežasties modelis daugiau dėmesio skirs didelių logotipų aptikimui, ignoruodamas mažų logotipų aptikimą. Todėl svarbu efektyviai pagerinti mažų logotipų aptikimo tikslumą skiriant dėmesį duomenų rinkiniuose esančių objektų dydžių balansui, tuo pat metu išlaikant didelių logotipų aptikimo tikslumą.

Automobilių logotipų aptikimo uždavinys – vienas iš aktyviai analizuojamų mažų logotipų aptikimo temų. Kompleksinio fono atveju, automobilio logotipas paprastai užima tik nedidelę vaizdo dalį, todėl tiksliai identifikuoti automobilio logotipus realiose situacijose yra keblu. Atsižvelgiant į dažną

greitaveikos reikalavimą, YOLO modeliai yra labai populiarūs automobilių logotipų aptikimo srityje [40] [41] [42].

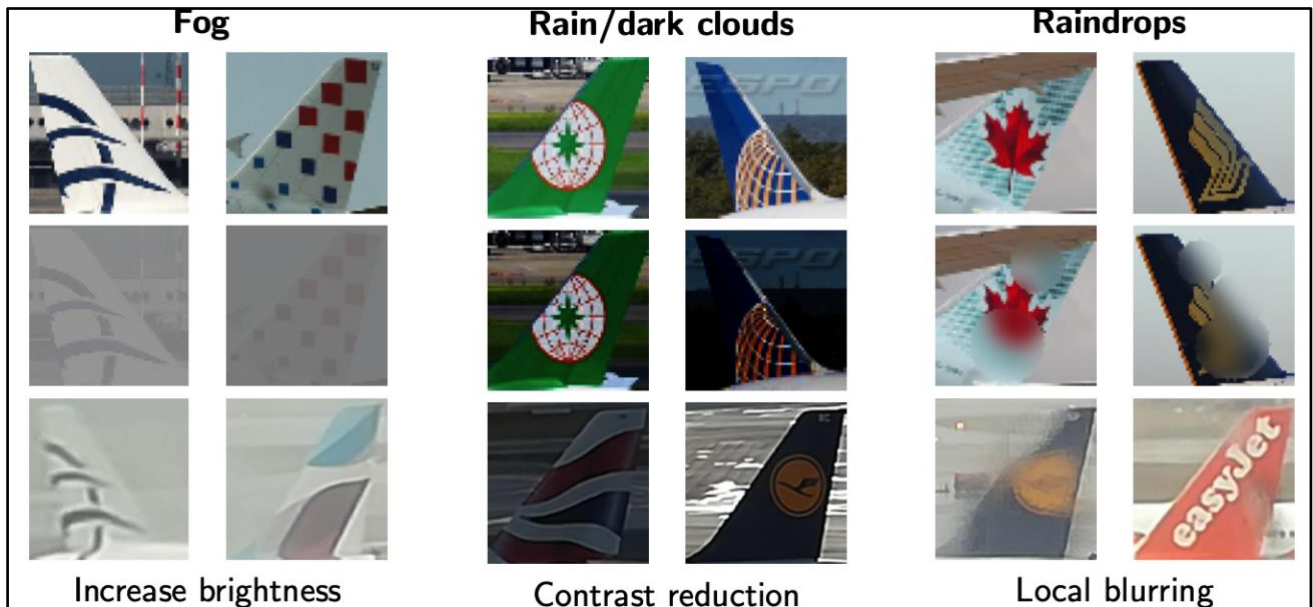
Kai kurie tyrėjai padidina mažų logotipų vaizdų skaičių naudodami duomenų augmentaciją, todėl modelis taip pat turi galimybę daugiau dėmesio skirti mažiems logotipo pavyzdžiams mokymo procese. Atliktame transporto priemonių logotipų aptikimo sudėtingose scenose tyrime [41], autoriai pasiūlė naują, duomenimis pagrįstą patobulintą mokymo metodą, panaudojant YOLOv3. Jie sujungė funkcijų ištraukimo tinklą su kelių masto sprendimų schema, kad pagerintų transporto priemonių logotipų aptikimo tikslumą. Tačiau transporto priemonių logotipų aptikimo sudėtingose scenose rezultatai vis dar išliko nepatenkinami. Remiantis [41] nuoroda, autoriai [42] išanalizavo transporto priemonės identifikavimo ypatybes statiniuose vaizduose ir nustatė, kad tiek funkcijų ištraukimo tinklas, tiek aptikimo mokymo strategija turi reikšmingos įtakos transporto priemonės identifikavimo tikslumui. Jie pasiūlė lengvą tinklo struktūrą su atskiriamais posūkiomis, kad pakeistų tradicinius metodus, siekiant pagerinti aptikimo tikslumą ir greitį pagal YOLO pagrįstą sistemą. Eksperimentai rodo, kad šis metodas pagerina transporto priemonės logotipo aptikimo realiuoju laiku našumą, taip pat pagerina nedidelio masto objektų aptikimo tikslumą, išlaikant greitį iki tam tikro lygio. Pridėjus periferinę informaciją aplink mažą logotipą kaip papildomus mokymo duomenis, galima gauti turtingesnę kontekstinę informaciją, kad būtų lengviau aptikti transporto priemonės logotipą [4].

### 1.7.2. Logotipų matomumo kompleksiskumas

Natūralios vaizdinės scenos paprastai yra gan įvairios ir neretai sudėtingos – logotipai turi keletą matomumo problemų, tokių kaip dalinis jų matomumas dėl kitų objektų persidengimo, formos pasikeitimas ir prastas perspektyvos efektas, o tai labai apsunkina logotipų aptikimą. Pastaraisiais metais akademiniai sluoksniai skyrė didelį dėmesį objektų aptikimui sudėtingose scenose. Kai vaizdą veikia nepalankūs veiksniai, tokie kaip apšvietimas, sukimasis, okliuzija ir pan., detektoriaus veikimas yra ženkliai prastesnis, todėl būtina sukurti efektyvų ir tvirtą logotipo detektorius [43] [44].

Atliktame oro linijų logotipo aptikimo realiomis oro sąlygomis tyrime [43], autoriai prasto matomumo problemą sprendžia taikydami objektų aptikimą su *AttentionMask* [45] – darbe pasiūloma logotipų aptikimo sistema nepalankiomis sąlygomis ir prastu matomumu (žr. **16 pav.**). Kadangi oro linijų logotipai ant lėktuvų yra gan dideli, *AttentionMask* sistema pašalina mažų objektų mastelį, kad pagerintų vykdymo efektyvumą ir sumažintų klaidingų teigiamų rezultatų skaičių. Tam, kad būtų imituojamas atšiaurių orų poveikis ir duomenų rinkinys būtų įvairesnis, autoriai darbe taiko duomenų papildymo sprendimą. Eksperimentiniai rezultatai rodo, kad siūloma strategija gali pasiekti gerą logotipo aptikimo našumą (mAP 0.69 ir mAP-75 naudojant YOLOv3 bei mAP 0.66 ir mAP-75 0.82 naudojant greitesnį R-CNN) sudėtingose realaus pasaulio aplinkose.





16 pav. Skirtingų oro sąlygų duomenų pavyzdžiai

Atsižvelgiant į tai, kad logotipų aptikimas sudėtinguose scenarijuose yra kompleksiškas dėl didelės logotipų tipų ir išvaizdos įvairovės, logotipų aptikimo uždaviniui spręsti darbo autoriai pasiūlė logotipo aptikimo sistemą, pavadintą *LogoNet*, kuri apima erdvinio dėmesio modulį [44]. Siūloma *LogoNet* sistema, lyginant su įprastais aptikimo algoritmais, sutelkia dėmesį į logotipo objektus vaizde kur kas tiksliau. Darbo eksperimentų rezultatai rodo, kad našumas, naudojant *FlickrLogos-32* duomenų imtį, pagerėjo maždaug 1,5 %, palyginti su kitais sprendimais.

*LogoNet* pasiekia reikšmingą našumo padidėjimą per ilgą skaičiavimo laiką. Norėdami aptikti tą patį logotipą su skirtingomis vaizdinėmis savybėmis, tyrime [46] šią problemą išsprendė išmokęs perkeliama ir lankstus kryžminio stiliaus santykio pagal meta-mokymosi strategiją. Be spartesnės RCNN sistemos, šiame metode yra ir brolio atšaka, skirta porų palyginimo aplinkai sukurti. Tuo tarpu detektoriaus klasifikavimo galvutė yra pertvarkyta į suderinimo modulį, kuris meta-mokosi, kaip klasifikuoti regionus per porinį suderinimą.

### 1.7.3. Duomenų rinkinių disbalansas

Siekiant atlikti „ilgos uodegos“<sup>1</sup> logotipų aptikimą, mašininio mokymosi modeliai yra apmokomi naudojant įvairius duomenų rinkinius, kuriuose yra ir įprasti, ir reti logotipai. Šie modeliai išmokomi atpažinti įvairias, su logotipais susijusias, vaizdines ypatybes ir raštus. Tačiau, duomenų rinkiniuose skirtingų logotipų pasiskirstymas dažnai būna netolygus. Naudojant šiuos duomenis modelio mokymui, „galvos“ (angl. *head*) duomenys dažniausiai yra per daug pritaikyti, tačiau „uodegos“ (angl. *tail*), dėl per mažo duomenų skaičiaus – per mažai [47] [48].

Šios problemos sprendimui moksliniame darbe siūlomas tvirtas bazinis metodas, pagrįstas YOLOv3 algoritmu, *Logo-Yolo* [16]. Pirmiausia, autoriai taiko *K-means* klasterizacijos algoritmą - pasirenkamas galimų inkaro dėžučių skaičius ir kraštinių santykio matmenys. Tada, panaudojama židinio paradigma, su kuria pašalinamas mėginių disbalansas duomenų rinkinyje. Galiausiai, panaudojama Clou

<sup>1</sup> Ilgos uodegos logotipų aptikimo problema - kai duomenų rinkiniuose nėra tolygaus duomenų pasiskirstymo - mažiau paplitusių arba rečiau sutinkamų logotipų yra mažiau, todėl jų mokymo tikslumas, lyginant su kitais logotipais, yra kur kas žemesnis.

paradigma, kuri pagerina ribinės dėžutės regresijos rezultatus. Lyginant su YOLOv3 algoritmu, šio metodo našumas žymiai pagerėjo numatant mažus objektus ir objektus sudėtingame fone, taip pat tiksliau aptikti retesni logotipai. Tačiau tam tikrais atvejais, metodas vis dar stokoja tikslumo – matome prastą našumą aptinkant panašius arba užkimštus logotipus.

#### **1.7.4. Atviro rinkinio logotipų aptikimo problema**

Šiuo metu, vis dar pačiais populiariausiais logotipų aptikimo metodais išlieka tie, kuriuose modeliai apmokomi aptikti iš anksto žinomas, konkrečius prekės ženklus. Sprendimai, kuriuose logotipų aptikimas neapsiriboja tik iš anksto žinomų logotipų aptikimu, vis dar yra pakankamai nauja tema, kuri patraukė tik nedaugelio tyrėjų dėmesį [14] [49] [50].

Ankstesniais tirtų logotipų aptikimo atvejais buvo daroma prielaida, kad kiekvienoje klasėje turime turėti daug mokymui skirtų duomenų, taip pat jiems priklausančių ribinių langelių anotacijų. Tačiau ši prielaida, sprendžiant atviro rinkinio logotipo aptikimo problemą, yra negaliojanti, nes logotipų klasės yra neapibrėžiamos ir iš anksto neįmanoma detalai žinoti kiekvieno logotipo vaizdo. Šiuo tikslu, problemos sprendimui autoriai pasiūlė atvirą logotipų aptikimo etaloną ir duomenų papildymo strategiją, kuri orientuota į logotipo konteksto optimizavimą: kontekstinę priešpriešinę mokymąsi (angl. *CAL – Contextual Adversarial Learning*) [14]. CAL paima logotipus kaip įvestis ir generuoja sintetinius, kontekstą atitinkančius, vaizdus. Šiuos naujai sugeneruotus paveikslėlius galima naudoti kaip papildomus mokymo duomenis, tačiau šio sintetinio vaizdo pasiskirstymas skiriasi nuo tikrojo bandomojo vaizdo. Norėdami išspręsti šią problemą, tie patys autoriai pasiūlė kelių perspektyvų kryžminės klasės (angl. *MPCC – Multi perspective Cross Class*) domenų pritaikymo metodą [49]. MPCC atlieka funkcijų paskirstymo suderinimą duomenų papildymo principu iš dviejų perspektyvų. Pirma – atlikti sintetinių logotipų vaizdų, kurie yra vieno karto piktogramų stebimos klasės, ir autentiškų logotipų vaizdų, kurie yra visiškai stebimos klasės, požymių pasiskirstymo suderinimą. Suderinus požymių pasiskirstymus tarp sintetinių ir autentiškų vaizdų, modelis gali efektyviai pasinaudoti sintetiniais duomenimis, kad pagerintų veikimą, ypač klasėms, turinčioms ribotą autentiškų mokymo duomenų imtį. Kitas būdas – suderinti logotipo ir ne logotipo vaizdų pasiskirstymą.

## 2. Reklamos efektyvumo vertinimo architektūra

Šiame skyriuje aprašoma kuriamos prekių ženklų reklamos efektyvumo vertinimo sistemos veikimo architektūra bei pateikiamos reikalavimų specifikacijos, funkciniai ir nefunkciniai reikalavimai bei funkcijų veikimo principai.

### 2.1. Sistemos panaudojimo atvejai

Šiame darbe siekiama sukurti mašininio mokymosi modelį, gebantį vaizdo įrašuose aptikti prekių ženklų reklamas, pateikti reklamos efektyvumo vertinimo rezultatus. Tam, kad būtų galima šį modelį panaudoti realiose situacijose, dėmesys skiriamas ir šio modelio eksploatacijai – sistemai, kuri gebėtų atlikti analizę ir gauti rezultatus iš bet kurio pasirinkto vaizdo įrašo be papildomo programavimo. Šis darbas siekia ne tik sukurti veiksmingą techninį sprendimą, bet ir užtikrinti, kad jis būtų lengvai panaudojamas ir praktiškai pritaikomas, suteikiantis galimybę naudotojams gauti reikalingą informaciją iš įvairių vaizdo turinio šaltinių panaudojant minimalias pastangas.

Kuriant prekių ženklų reklamos efektyvumo vertinimo sistemą, numatyti du panaudos atvejai: atlikti vaizdo įrašo analizę bei apdoroti reklamos efektyvumo vertinimo duomenis. (žr. **17 pav.**). Kiekvienas atvejis detalizuojamas lentelėse (žr. **5 lentelė** – **6 lentelė**).



**17 pav.** Panaudos atvejų diagrama

#### 5 lentelė. „Atlikti vaizdo įrašo analizę“ panaudos atvejo aprašas

Panaudojimo atvejis	Atlikti vaizdo įrašo analizę
Tikslas	Atlikti pasirinkto vaizdo įrašo analizę
Aktoriai	Naudotojas
Prieš – sąlygos	Kompiuteryje išsaugotas analizei norimas panaudoti vaizdo įrašas.
Sužadavimo sąlyga	Sistemos naudotojas nori atlikti vaizdo įrašo analizę.
Po – sąlyga	Gaunami vaizdo įrašo analizės rezultatai.
Pagrindinis scenarijus	<ol style="list-style-type: none"> <li>1. Kompiuteryje pasirenkamas analizei skirtas vaizdo įrašas.</li> <li>2. Vaizdo įrašas padorojamas algoritmais.</li> <li>3. Išsaugomi gauti rezultatai.</li> </ol>
Alternatyvūs scenarijai	–

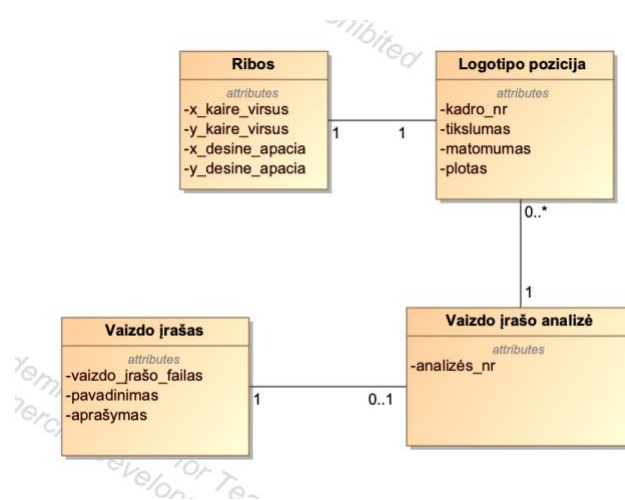
#### 6 lentelė. „Apdoroti reklamos efektyvumo vertinimo duomenis“ panaudos atvejo aprašas

Panaudojimo atvejis	Apdoroti reklamos efektyvumo vertinimo duomenis
Tikslas	Susisteminti analizės metu gautus duomenis, atvaizduoti juos grafikuose ir pateikti išvadas.
Aktoriai	Naudotojas
Prieš – sąlygos	Sistema paruošta naudoti. Gauti vaizdo įrašo analizės rezultatai.
Sužadinimo sąlyga	Sistemos naudotojas nori apdoroti pasirinkto vaizdo įrašo analizės rezultatus.
Po – sąlyga	Išsaugomas apdorotų duomenų rezultatas.
Pagrindinis scenarijus	<ol style="list-style-type: none"> <li>1. Sistemos naudotojas pasirenka vaizdo įrašo analizės metu gautą rezultatų failą.</li> <li>2. Duomenys apdorojami algoritmais.</li> <li>3. Išsaugomas apdorotų duomenų rezultatas.</li> </ol>
Alternatyvūs scenarijai	–

## 2.2. Sistemos statinis vaizdas

Sistemą sudaro keturios klasės (žr. **18 pav.**):

- vaizdo įrašo klasė, kuri saugo vaizdo įrašo failą, pavadinimą bei aprašymą;
- vaizdo įrašo analizės klasė, kurią sudaro analizės numeris. Vaizdo įrašo analizė gali priklausyti tiek vienam konkrečiam vaizdo įrašui;
- logotipų pozicijos klasė, kurią sudaro kadro numeris, tikslumas, matomumas bei užimamas plotas. Vaizdo įrašo analizei priklauso daug logotipo pozicijų;
- ribos klasė, kurią sudaro viršutinio kairiojo ir apatinio dešiniojo aptikti logotipo ribos koordinatės.



**18 pav.** Vaizdo įrašų analizės sistemos klasių diagrama

## 2.3. Funkciniai ir nefunkciniai reikalavimai

Išsikelti funkciniai reikalavimai:

- sistema, naudojanti kompiuterinės regos technologijas ir mašininio mokymosi algoritmus, aptinka ir lokalizuoja prekės ženklus, jų reklamas vaizdo įrašo kadruose;

- sistema, atsižvelgdama į kameros judesius ir perspektyvos pokyčius, geba sekti prekės ženklų reklamų judėjimą;
- sistema pateikia aptiktas prekės ženklų reklamas ir suteikia apie jas informaciją;
- sistema, taikydama reklamos efektyvumo vertinimo formules, pateikia naudotojui apdorotus analizės duomenų rezultatus.

Išsikelti nefunkciniai reikalavimai:

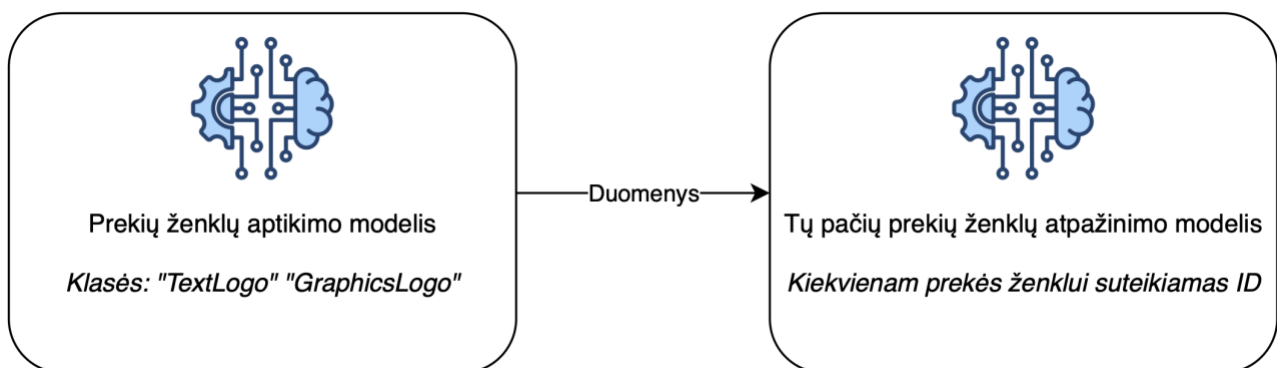
- sistema siekia aukštą tikslumo lygį, aptikdama ir sekdamą prekės ženklų logotipus įvairiomis apšvietimo sąlygomis, įvairias kameros pasukimo kampais ir skirtingais fonais (aukštas tikslumas laikomas virš 80%);
- sistema palaiko didelį duomenų srautą (1080p 60 kadrų per sekundę vaizdo įrašus);
- sistema suderinama su skirtingais populiariausiais vaizdo įrašų formatais, dažniausiai naudojamais tiesioginiuose sporto vaizdo įrašuose, užtikrinant sklandų integravimą su skirtingais šaltiniais ir sumažinant suderinamumo problemas.

## 2.4. Sistemos veikimo detalizavimas

Šiame skyriuje aprašomas kuriamos sistemos veikimas, detalizuojami funkciniai reikalavimų veikimo principai ir veiksmų sekos.

### 2.4.1. Bendras kuriamos sistemos aprašas

Šiame darbe pasiūlomas naujas prekių ženklų bei jų reklamų aptikimo sprendimas, kuris geba vaizdo įrašuose aptikti visus prekės ženklus bei jų reklamas. Šis rezultatas pasiekiamas sukuriant dvi giliojo mokymosi grįstas sistemos dalis (žr. **19 pav.**).



**19 pav.** Darbe sukurtų mašininio mokymosi dalių tarpusavio komunikacija

Prekės ženklų aptikimui sukurtas sprendimas, kuris aptikimo procese neprisiriša prie klasės, kuri nurodo būtent prekės ženklo pavadinimą. Vietoj to, prekės ženklai aptinkami panaudojant dvi bendrines klases – tekstiniai prekės ženklai bei grafiniai prekės ženklai. Tokiam sprendimui įgyvendinti, prieš mokymą reikalingas papildomas duomenų paruošimas, kurio metu turi būti pakeičiamos visų prekės ženklų klasės. Tai atliekama pasitelkiant giliojo mokymosi teksto atpažinimo modelį CRAFT.

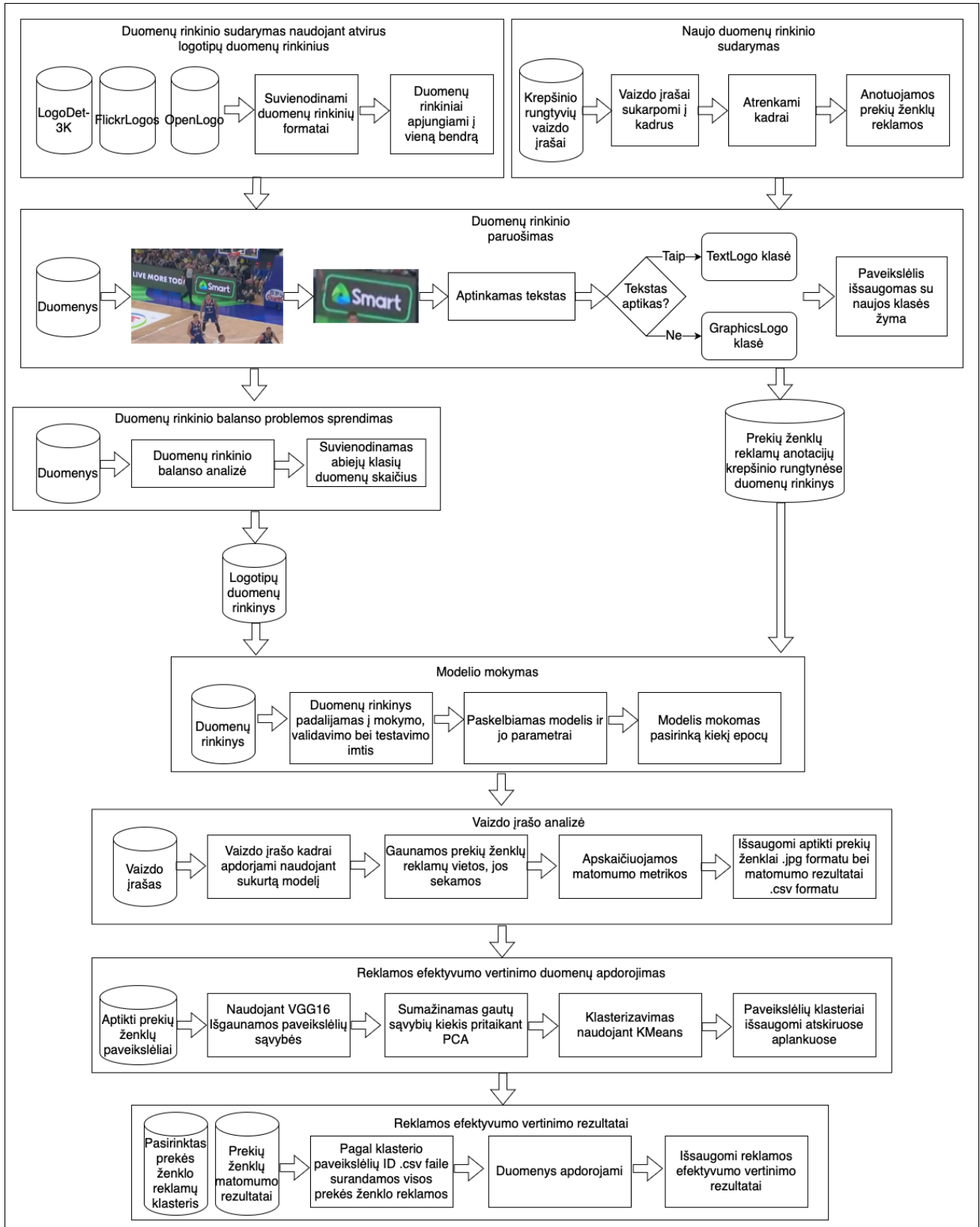
Šiuo metodu įgyvendintas sprendimas turi minusą - aptikti prekės ženklai netenka savo prekės ženklo pavadinimo klasės, todėl analizės metu tampa neaišku, kurie prekės ženklai yra skirtingi, kurie tie

patys. Šios problemos sprendimui pasitelkiamas antras mašininio mokymosi modelis, kuris suskirsto aptiktus prekės ženklus pagal vaizdines savybes į atskirus klasterius. Šių klasterių dėka, galime nesunkiai identifikuoti tuos pačius prekės ženklus ir analizuoti reklamos efektyvumo vertinimo rezultatus.

Sistemą sudaro šie etapai:

1. duomenų rinkinio sudarymo etapas, kurio metu, panaudojant pačius populiariausius ir didžiausius atvirus duomenų rinkinius, sudaromas vienas bendras modelio mokymui skirtas duomenų rinkinys;
2. naujo duomenų rinkinio sudarymo etapas, kurio metu panaudojami krepšinio rungtynių kadrai ir juose anotuojami visi matomi prekės ženklai ir jų reklamos;
3. duomenų rinkinio paruošimo mokymui etapas, kurio metu, pasitelkiant teksto aptikimo mašininio mokymosi modelį, duomenų rinkinyje klasės pakeičiamas į dvi naujas klases – tekstinių prekės ženklų ir ne tekstinių prekės ženklų klases;
4. duomenų rinkinio balanso etapas, kurio metu atliekama duomenų rinkinio analizė, identifikuojamos ir sprendžiamos balanso problemos;
5. modelio mokymo etapas, kurio metu paskelbiamas modelis, pasirenkami jo parametrai bei naudojant sudarytus duomenų rinkinius, mokomas aptikti prekės ženklus pasirinktą kiekį epochų;
6. vaizdo įrašo analizės etapas, kurio metu pasirinktas vaizdo įrašas apdorojamas mašininio mokymosi modeliu bei gauti rezultatai išsaugomi atitinkamais formatais;
7. reklamos efektyvumo vertinimo duomenų apdorojimo etapas, kurio metu apdorojami gauti aptikimo rezultatai – prekių ženklai klasterizuojami pagal vaizdines savybes;
8. reklamos efektyvumo vertinimo etapas, kurio metu apdorojamas pasirinkto prekės ženklo klasteris bei pateikiami jo susisteminti reklamos efektyvumo vertinimo rezultatai.

Šių etapų detalesnis veikimas pavaizduotas schemoje (žr. **20 pav.**).

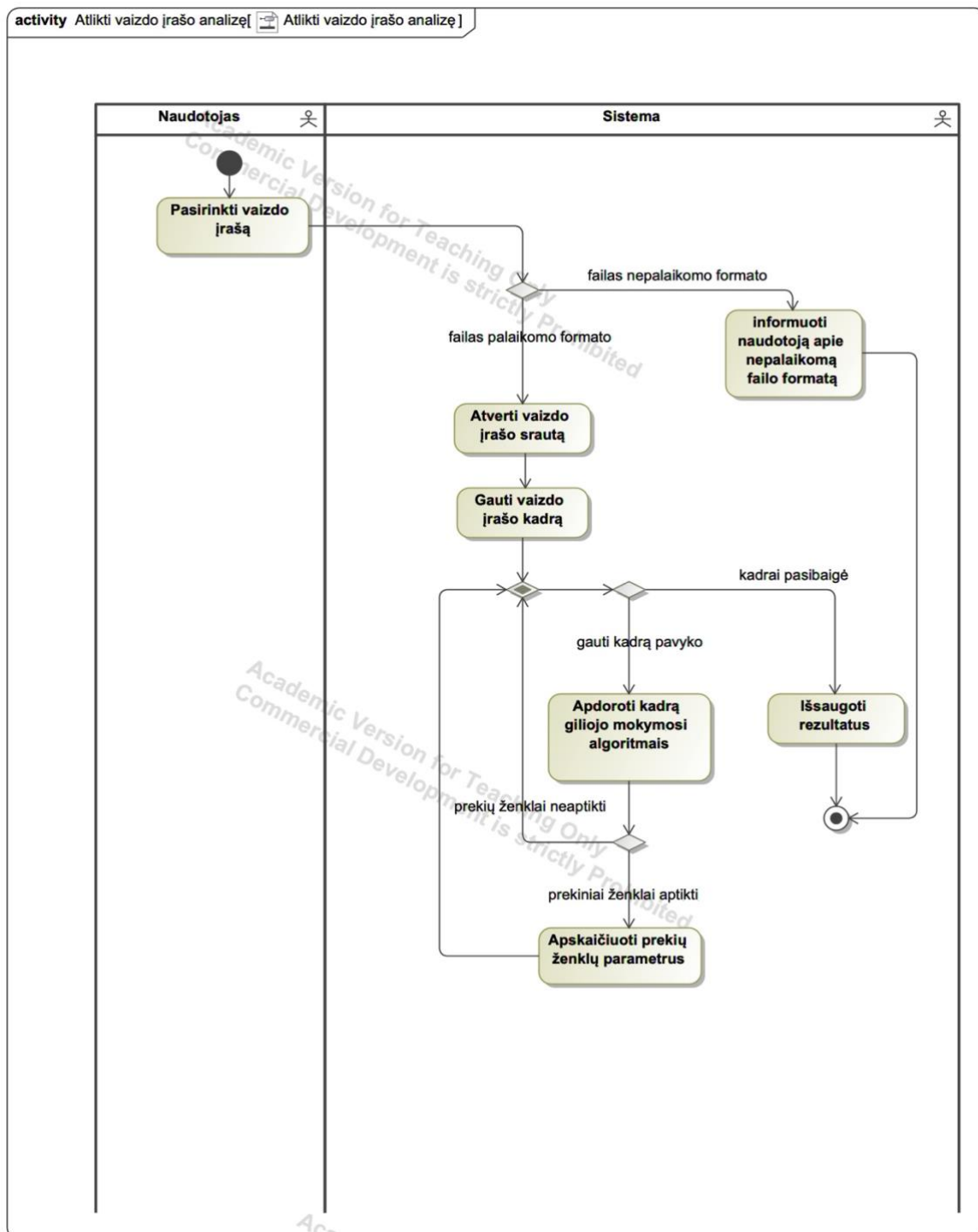


20 pav. Sukurtos sistemos veikimo principo schema

## 2.4.2. Vaizdo įrašo apdorojimo panaudos atvejo aprašas

Vaizdo įrašo analizės kartinė detalė – vaizdo įrašas. Sistemoje palaikomi vaizdo įrašų formatai – *mov*, *avi*, *mp4*, *mpg*, *mpeg*. Vaizdo įrašas turi būti išsaugotas kompiuteryje.

Pasirinktus tinkamo formato vaizdo įrašą, sistema pradeda jo analizę. Vaizdo įrašai yra padalijami į kadrus, kur kiekvienas kadras yra atskirai apdorojamas. Priklausomai nuo vaizdo įrašo ilgio, kadru per sekundę skaičiaus bei raiškos, analizė gali šiek tiek užtrukti. Pasibaigus analizei, rezultatai išsaugomi lokacijoje, kur yra patalpintas vaizdo įrašas (žr. **21 pav.**)

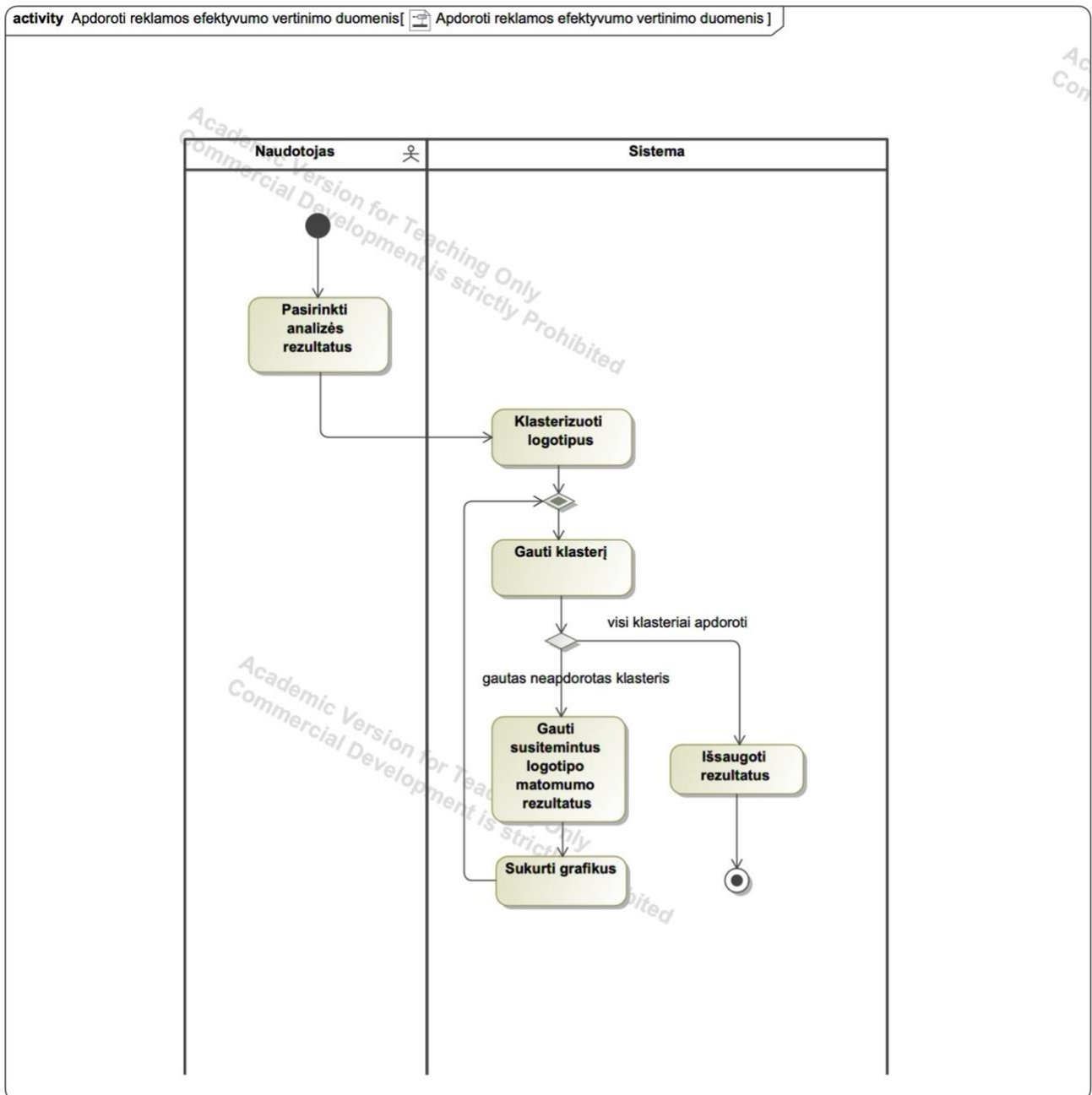


21 pav. „Atlikti vaizdo įrašo analizę“ veiklos diagrama



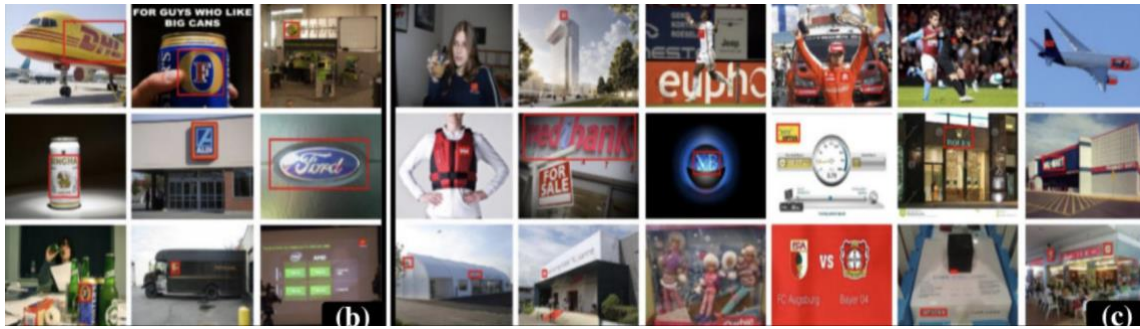
### 2.4.3. Reklamos efektyvumo vertinimo duomenų apdorojimo panaudos atvejo aprašas

Kai naudotojas sėkmingai atlieka vaizdo įrašo analizę ir gauna duomenis apie kadruose matomus prekės ženklus, galima pradėti rezultatų sisteminimą. Reklamos efektyvumo vertinimo duomenų apdorojimo eiga atvaizduota veiklos diagramoje (žr. **22 pav.**). Sistema suskirsto visus logotipus į klasterius pagal vaizdines jų savybes. Kiekvienas klasteris apdorojamas atskirai – apskaičiuojamos klasteriui priklausančių logotipų matomumo savybės.



22 pav. „Apdoroti reklamos efektyvumo vertinimo duomenis“ veiklos diagrama





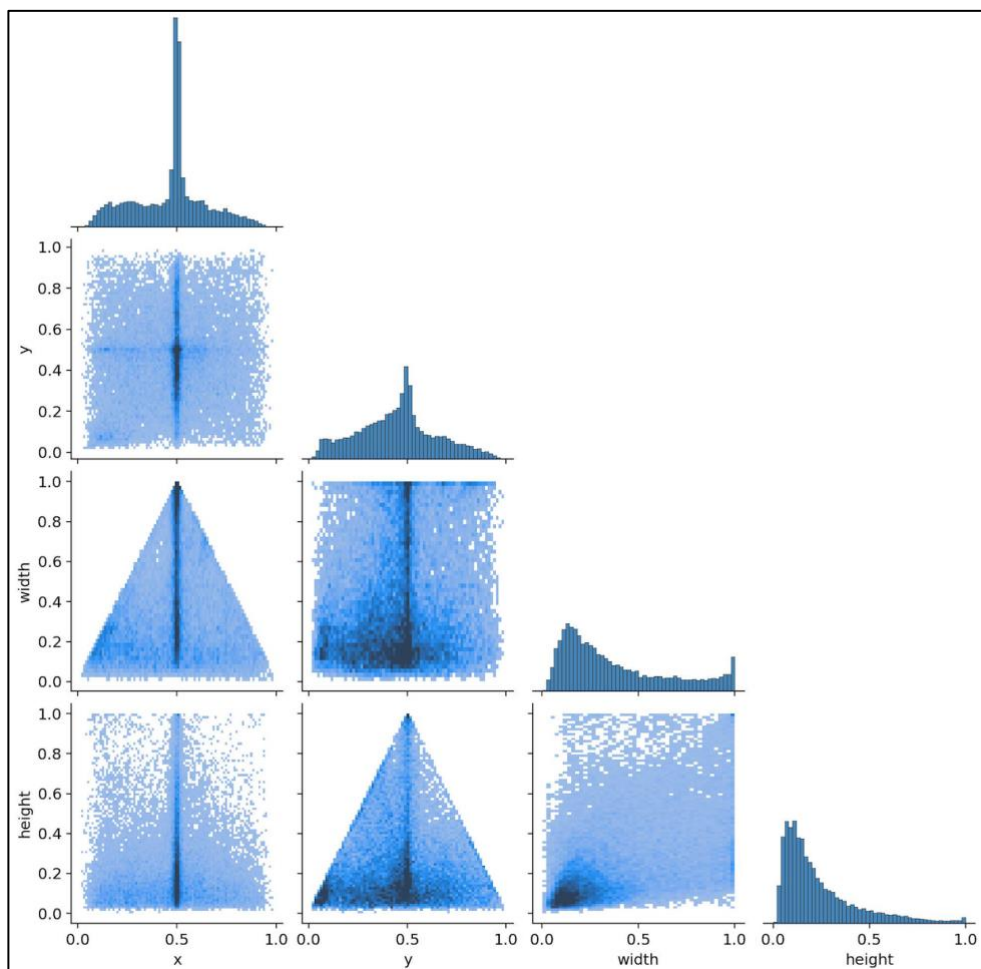
25 pav. SportLogo duomenų peržiūra

Prieš pradėdant kurti mašininio mokymosi modelius, svarbu atlikti keletą techninių analizių, kurių tikslas – įvertinti duomenų kokybę ir tinkamumą. Norint gauti įžvalgų apie duomenų rinkinio sudėtį ir charakteristikas, būtina taikyti tiriamosios duomenų analizės (EDA) metodus. Tai apima logotipų kategorijų paskirstymo apskaičiavimą, skirtingų logotipo atributų (formų dydžių, teksto) dažnumo tyrimą ir bet kokio klasių disbalanso ar paklaidų, galinčių turėti įtakos tolesnėms ML užduotims, nustatymą. Pasirinktų duomenų imčių aprašymai pateikti lentelėse (7 lentelė – 8 lentelė).

7 lentelė. FlickLogos duomenų rinkinio aprašas

Duomenų rinkinio dalis	Aprašymas	Paveikslėliai	Viso paveikslėlių
P1	Atrinkti paveikslėliai su švairiu fonu. Kiekviename paveikslėlyje po vieną prekės ženklo logotipą.	10 kiekvieno logotipo	320
P2	Skirtingi paveikslėliai kuriuose matomas bent vienas prekės ženklo logotipas įvairiomis sąlygomis. Paveikslėliai be logotipų.	30 kiekvieno logotipo 3000 paveikslėlių be logotipų	3960
P3	Skirtingi paveikslėliai kuriuose matomas bent vienas prekės ženklo logotipas įvairiomis sąlygomis. Paveikslėliai be logotipų.	30 kiekvieno logotipo 3000 paveikslėlių be logotipų	3960

Atlikta *LogoDet-3K* duomenų rinkinio žymų analizė atvaizduota poriniame grafike (dar vadinamoje sklaidos diagramų matricoje) (žr. 26 pav.). Histogramose galime pastebėti, kad x pasiskirstymas atrodo gana tolygus su nedidele koncentracija ties centru, o y pasiskirstymas rodo stiprią centravimo tendenciją – dauguma verčių susitelkia aplink vidurį. Sklaidos diagramose matomas centrinis x ir y verčių susitelkimas, o plotis ir aukštis indikuoja trikampio pasiskirstymą – tai nurodo, kad ekstremalių verčių yra nedaug. Ryšys tarp pločio ir aukščio rodo teigiamą koreliaciją, kur didesni plotis dažnai atitinka didesnę aukštį, tačiau didesnės vertės rodo didesnį kintamumą. Ši analizė parodo ir tam tikrą centrinį susitelkimą ir duomenų simetriją, ypač aplink x ir y vidurius.



26 pav. LogoDet-3K duomenų rinkinio analizė

8 lentelė. LogoDet-3k duomenų rinkinio aprašas

Kategorija	Paveikslėlių skaičius	Prekių ženklų skaičius
Maisto prekių ženklai	53,350	64,276
Drabužių prekių ženklai	31,266	37,601
Būtinybių prekių ženklai	24,822	30,643
Kiti prekių ženklai	15,513	20,016
Elektronikos prekių ženklai	9,675	12,139
Transporto prekių ženklai	10,445	12,791
Laisvalaikio prekių ženklai	5,685	6,573
Sporto prekių ženklai	3,953	5,041
Medicinos prekių ženklai	3,945	5,185
Viso	158,652	194,261

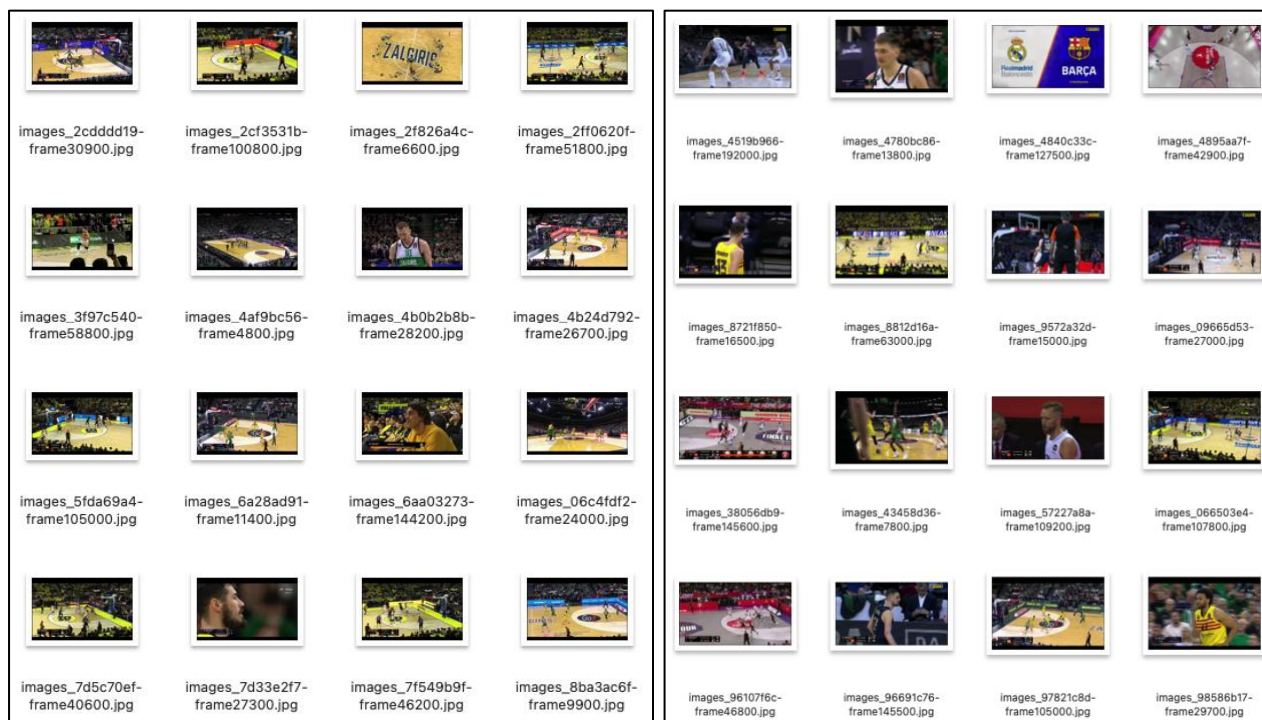
### 3.2. Duomenų rinkinio kūrimas

Tolimesniam modelių testavimui realiose situacijose, sukurtas naujas duomenų rinkinys. Kuriant duomenų rinkinį, atsižvelgta į šiuos scenarijus:

- skirtingi kamerų kampai – kadrai iš arti, platūs kadrai, kadrai iš toli, šoninių kamerų kadrai;

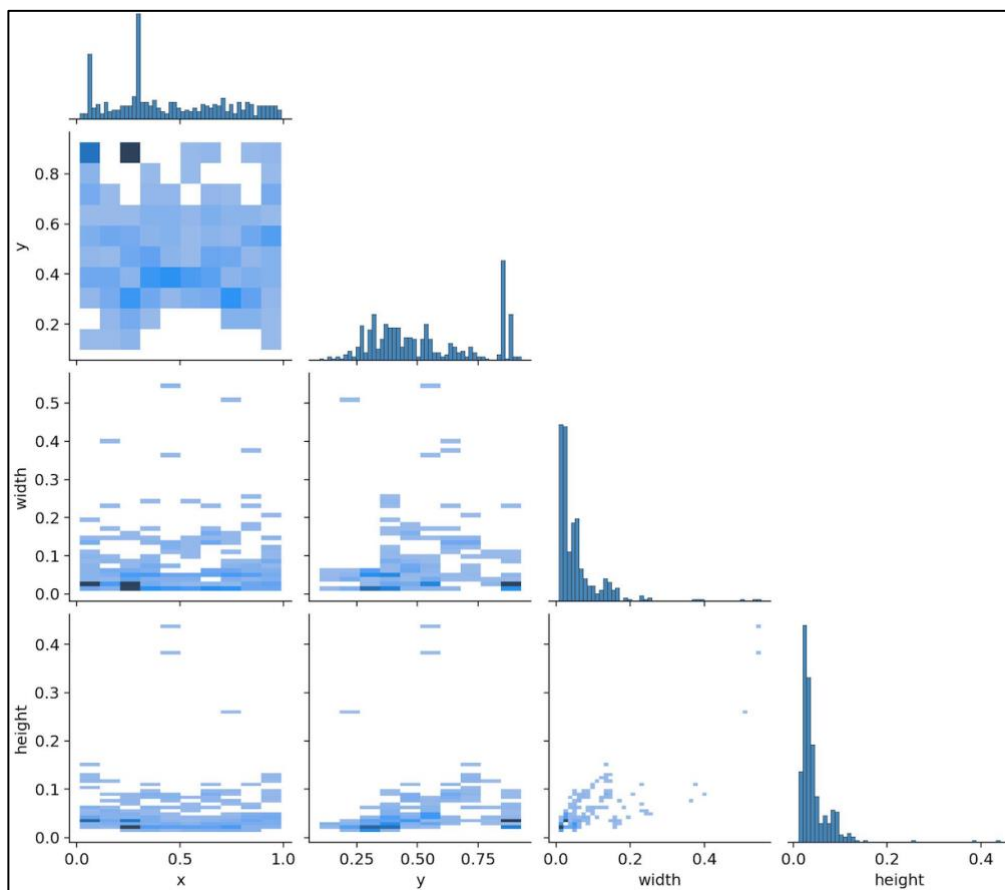
- kadrai su skirtingomis apšvietimo sąlygomis;
- įvairios prekių ženklų pozicijos;
- susilieję kadrai – greito judesio kadrai, kuriuose objektai matomi neryškiai;
- persidengiantys objektai – prekių ženklai iš dalies dengiami žaidėjų ar kitų objektų;
- įvairių rezoliucijų kadrai.

Sukurta duomenų rinkinį sudaro 1500 kadru iš 10 skirtingų krepšinio varžybų įrašų (žr. **27 pav.**).



**27 pav.** Sudaryto duomenų rinkinio peržiūra

Atliktoje duomenų rinkinio analizėje, kuri atvaizduota poriniame grafike (žr. **28 pav.**), pateikiama išsami duomenų rinkinio ribojančių langelių koordinacių ir dydžių pasiskirstymo ir ryšių matrica. Histogramos ir sklaidos diagramos rodo, kad dauguma ribojančių langelių yra sąlyginai mažo dydžio, tačiau egzistuoja keletas objektų, kurie išsiskiria savo dydžiu iš kitų. Taip pat galime pastebėti, kad padėtys yra gana įvairios, nėra stipriai koreliuojančios.



28 pav. Sukurto prekės ženklų krepšinio rungtyne duomenų rinkinio analizė

### 3.3. Duomenų rinkinių apdorojimas

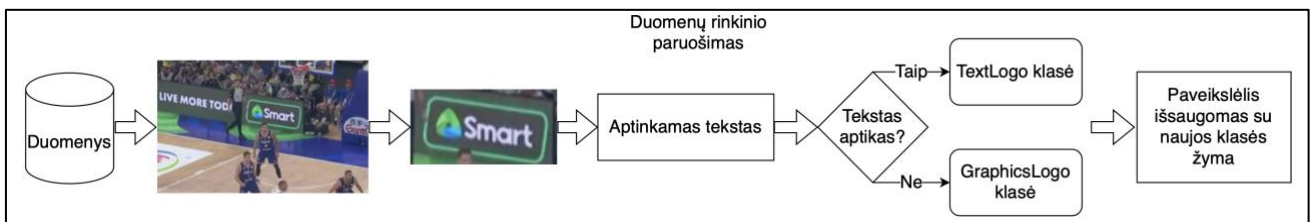
Tam, kad standartinis objektų aptikimo modelis veiktų, reikia bent dviejų skirtingų aptinkamų objektų klasių. Logotipų aptikimo duomenų rinkiniuose kiekvienas prekės ženklas yra priskiriamas kategorijai pagal prekės ženklo pavadinimą. Vadinasi, duomenų rinkiniai turi tiek klasių, kiek juose yra skirtingų prekių ženklų. Toks duomenų rinkinio principas turi trūkumų:

- modelis turi būti apmokomas aptikti kiekvieną prekės ženklą atskirai. Tam reikalingas didelis duomenų kiekis su kiekvienu prekės ženklu;
- duomenų rinkinys turi būti subalansuotas;
- modelį reikia kaskart iš naujo mokyti aptikti žinomus, bet išvaizdą pakeitusius prekių ženklus;
- reikia sekti kiekvieną naują prekės ženklą. Tai lemia, kad modelis turi būti kaskart vis iš naujo apmokomas.

Šiame darbe siekiama sukurti visų prekių ženklų globalų aptikimo modelį, kuris gebėtų aptikti tiek senus, tiek naujus, tiek žinomus, tiek naujus prekių ženklus bei jų reklamas. Todėl duomenų rinkinys yra modifikuojamas – duomenys yra suskirstomi į **dvi klases** – **tekstinius prekių ženklus ir ne tekstinius prekių ženklus**. Duomenų rinkinio korekcija atliekama šiais žingsniais (žr. 29 pav.):

1. išskiriamos ribojamojo langelio koordinatės iš jau turimų anototų logotipų duomenų rinkinių;
2. vaizdas apkarpomamas pagal ribojamojo langelio koordinatas, taip gaunamas konkretus prekės ženklo regionas;

3. apkarpytas vaizdas apdorojamas naudojant CRAFT (angl. *Character-Region Awareness For Text*) teksto aptikimo modelį;
4. jei aptinkamas tekstas, pažymime tą konkretų ribojimo laukelį kaip „TextLogo“. Kitu atveju apribojantį langelį pažymime kaip „GraphicsLogo“.



29 pav. Duomenų rinkinio apdorojimo eiga

Apdorotų atvirų logotipų duomenų rinkinių klasių pasiskirstymo rezultatai pavaizduoti lentelėje (žr. **9 lentelė**). Galime pastebėti labai didelį duomenų rinkinio disbalansą – tekstinių prekės ženklų yra gerokai daugiau nei grafinių. Dėl netolygaus duomenų pasiskirstymo, modelis bus šališkas tekstiniams prekių ženklams, dėl to nukentės klasifikavimo rezultatai testavimo metu. Klasių balanso problema sprendžiama panaudojant vienodą kiekį abiejų klasių duomenų – visus grafinius prekės ženklus bei atitinkamą dalį įvairių, atsitiktine tvarka parinktų, tekstinių prekių ženklų. Galutinį duomenų rinkinį sudaro 27 796 grafiniai prekės ženklai bei 27 796 tekstiniai prekės ženklai.

Šiame darbe sukurtame krepšinio duomenų rinkinyje su anotuotais prekių ženklais bei jų reklamomis, klasių disbalanso problema nepastebėta – prekių ženklai pasiskirstę tolygiai. Kiekviename kadre pastebime po 6 – 8 grafinius prekės ženklus bei 7 – 10 tekstinių prekės ženklų.

**9 lentelė.** Duomenų rinkinių klasių pasiskirstymas po apdorojimo

Duomenų rinkinio pavadinimas	TextLogo klasė	Graphics Logo klasė
LogoDet-3K	168 416	25 845
FlickrLogos-32	4 910	734
SportLogo	1 619	1 217
Viso	174 945	27 796

Galutinis, modelio mokymui skirtas, duomenų rinkinys saugomas:

- COCO (angl. *Common Objects in Context*) formatu – tai duomenų formatas, kuris plačiai naudojamas objektams aptikti, segmentavimo modeliams mokyti ir vertinti.
- YOLO formatu.

COCO duomenų rinkinį sudaro:

- anotacijos: COCO duomenų formato pagrindinė dalis yra anotacijos informacija, kuri aprašo objektus ir jų savybes vaizduose. Kiekviena anotacija apima šią informaciją:
  - vaizdo ID: unikalus identifikatorius vaizdui, kuriam priklauso anotacija.
  - kategorijos ID: unikalus objekto kategorijos ar klasės identifikatorius.
  - ribojamasis langelis: koordinatės apibūdinančios stačiakampio formos žymą, kuri apima objektą vaizde. Šios koordinatės paprastai pateikiamos kaip (x, y, plotis, aukštis).
- vaizdai: taip pat pateikiama informacija apie vaizdus duomenų rinkinyje, įskaitant:
  - vaizdo ID: unikalus identifikatorius vaizdui.

- failo pavadinimas: vaizdo failo pavadinimas arba kelias.
- plotis ir aukštis: vaizdo matmenys taškais.
- kategorijos: COCO duomenų rinkiniai apima įvairias objektų kategorijas. Kiekvienai kategorijai priskiriamas unikalus kategorijos ID ir kategorijos pavadinimas.

YOLO duomenų rinkinį sudaro:

- anotacijos: tekstiniai failai tokiu pačiu pavadinimu kaip ir vaizdas, bet su plėtiniu .txt, kuriuose aprašomi ribojantys langeliai aplink objektus ir jų klases. Kiekviena komentarų failo eilutė žymi vieną vaizdo objektą;
- paveikslėlių ir anotacijų aplankai, kuriuose atskirai saugomi vaizdai bei anotacijos;
- mokymo, validacijos bei testavimo aplankai paveikslėlių bei anotacijų aplankuose, kuriuose saugomos duomenų rinkinių dalys.

Prieš pradėdant kurti mašininio mokymosi modelius, prireikia atlikti paveikslėlių apdorojimą. Atsižvelgiant į vaizdo dydžius, skiriamąją gebą ir spalvų erdvių skirtumus, atliekamos tam tikras operacijos:

- normalizavimas: normalizuojant vaizdus, pikselių reikšmės nustatomas iki standartizuoto diapazono. Tai padeda sumažinti pikselių intensyvumo svyravimų poveikį ir užtikrina, kad visų vaizdų duomenų pasiskirstymas būtų panašus;
- sukimas ir apvertimas: šios operacijos apima vaizdo pasukimą tam tikru kampu arba apvertimą horizontaliai arba vertikalčiai. Šios transformacijos padeda išplėsti duomenų rinkinį ir padidinti jo įvairovę, todėl modeliai būna atsparesni objektų padėties skirtumams;
- histogramos išlyginimas: histogramos išlyginimas, perskirstydamas pikselių intensyvumą, padidina vaizdo kontrastą. Tai padeda pagerinti detalių ir struktūrų matomumą, ypač esant prastam arba netolygiam apšvietimui;
- triukšmo šalinimas: vaizdo triukšmo mažinimo metodais siekiama sumažinti vaizduose esantį triukšmą arba artefaktus. Taikomas Gauso filtravimas, kuris padeda pagerinti vaizdo aiškumą ir padidinti gautų rezultatų tikslumą.

Funkcijų ištraukimas – taip pat labai svarbus duomenų apdorojimo žingsnis. Šis procesas apima neapdorotų vaizdo duomenų pavertimą prasmingu vaizdu, kurį gali suprasti ML algoritmai. Funkcijų ištraukimui taikomi metodai:

- nekintamo masto funkcijų transformavimas (angl. *Scale-Invariant Feature Transform – SIFT*): SIFT yra patikimas ir plačiai naudojamas funkcijų išgavimo metodas, kuris nustato pagrindinius taškus ir apskaičiuoja deskriptorius pagal vietinius vaizdo gradientus;
- orientuotų gradientų histograma (angl. *Histogram of Oriented Gradients - HOG*): HOG apskaičiuoja orientuotų gradientų histogramas vietiniuose vaizdo regionuose. Jis fiksuoja gradiento orientacijų pasiskirstymą.

Be to, labai svarbu padalyti duomenų rinkinį į tinkamus mokymo, validavimo ir testavimo dalis. Mokymo rinkinys naudojamas modelio mokymui, validavimo rinkinys naudojamas modelio validacijai ir hiperparametrų derinimui, o testavimo rinkinys yra naudojamas galutinio modelio veikimo įvertinimui. Duomenų paskirstymas: 80% mokymo duomenų imtis, 10% validavimo duomenų imtis, 10% testavimo duomenų imtis.



#### 4. Reklamos efektyvumo vertinimas

Reklamos efektyvumo vertinimas atliekamas remiantis SMV formule:

$$SMV = FMV \times MVP \times DF,$$

- FMV (angl. *Full Media Value*) – šią reikšmę nustato reklamos užsakymo vykdytojas;
- MVP (angl. *Media Value Percentage*) – ši reikšmė susideda iš prekės ženklo matomumo kokybės rodiklių, tokių kaip dydis, matymo kampas, matomos reklamos vieta ekrane ir pan.;
- DF (angl. *Duration Factory*) – šią reikšmę nusako kiek laiko prekės ženklas buvo matomas eteryje.

Padauginus šiuos tris komponentus kartu, SMV formulė suteikia išsamią metriką reklaminės kampanijos veiksmingumui įvertinti. Tai leidžia reklamuotojams ir rinkodaros specialistams įvertinti jų reklaminių pastangų sukuriama vertę, atsižvelgiant į finansines investicijas (FMV) ir kokybinius bei laikinus prekės ženklo eksponavimo aspektus (MVP ir DF). Šis daugialypis požiūris leidžia suprasti reklamos iniciatyvų poveikį ir padeda tobulinti būsimas strategijas siekiant optimalių rezultatų.

Tam, kad apskaičiuoti šiuos rodiklius reikia:

- apskaičiuoti, kiek kiekvienas prekės ženklas, kaskart pasirodęs ekrane, užima vietos. Apskaičiuojama reklamos užimama vieta (jeigu matomos keli tie patys prekių ženklai – plotas sumuojamas) (žr. **30 pav.**);



**30 pav.** Prekių ženklų užimamų plotų vizualizavimas

- apskaičiuoti, kokią poziciją kiekvienas prekės ženklas užima ekrane (kuo arčiau ekrano centro, tuo prekės ženklas geriau pastebimas). Apskaičiuojamas matomo prekės ženklo centro atstumas iki ekrano centro (žr. **31 pav.**);



**31 pav.** Prekių ženklų atstumo iki centro vizualizavimas

- apskaičiuoti, kiek kartų transliacijos metu buvo matomas kiekvienas prekės ženklas (prekės ženklas laikomas matomu, kai ekrane jis išlieka 1 sekundę). Jeigu vienu metu matomi keli tie patys prekių ženklai, laikoma, kad jis matomas tik vieną kartą (žr. **32 pav.**).



**32 pav.** Kelių tų pačių prekių ženklų akcentavimas

- apskaičiuoti santykį tarp analizuojamo matomo prekės ženklo užimamo reklamos ploto bei visų kitų prekių ženklų reklamų užimamo ploto. Kuo daugiau reklamų matoma tuo pačiu metu, tuo mažesnė tiriama prekės ženklo matomumo tikimybė.
- Apskaičiuoti matomos reklamos anamorfinę projekciją. Kuo aptiktas prekės ženklas labiau iškraipomas, tuo jo matomumas bei atpažinimas būna prastesnis.

## 5. Prekių ženklų aptikimo bei atpažinimo tiriamoji dalis

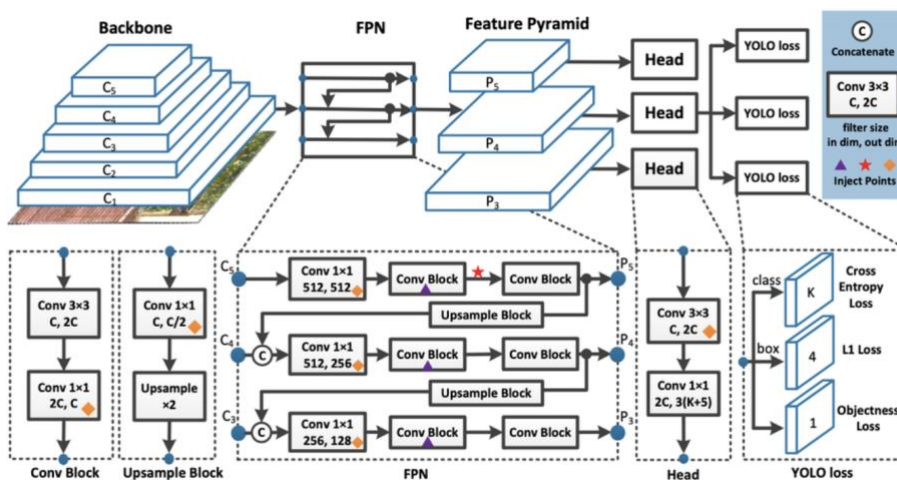
Tiriamosios dalies tikslai:

- įvertinti kiekvieno pasirinkto giliojo mokymosi algoritmo tikslumo rezultatus ir juos tarpusavyje palyginti. Rezultatų vertinimas atliekamas remiantis skirtingomis vertinimo metrikomis, tokiomis kaip atkūrimas, tikslumas, vidutinis bendras tikslumas, persidengimo rodikliai ir pan.
- iširti ir įvertinti algoritmų greitaveiką. Šios problemos uždavinys orientuotas į realaus laiko duomenų apdorojimą, todėl greitaveika vertinimas ypač svarbus.

### 5.1. Tiriamosios dalies aprašymas

Tiriant objektų aptikimo būdus, kuriuos būtų galima pritaikyti prekių ženklų reklamos aptikimui, nuspręsta palyginti vieno ir dviejų sluoksnių objektų aptikimo architektūras. Vieno sluoksnio metodas naudoja paprastą, seklią tinklo architektūrą, kuri apdoroja įvesties vaizdą vienu žingsniu. Šiuo metodu įgyvendinti sprendimai demonstruoja greitą objektų atlikimo laiką, nes tai reikalauja (lyginant su dviejų sluoksnių architektūra) kur kas mažesnių skaičiavimų kaštų, todėl jis būna tinkamesnis realiojo laiko programoms, kuriose greitis yra labai svarbus. Tačiau pastebima, kad modeliai neretai praranda tikslumą, kai aptinkami objektai yra iš dalies uždengti arba prastai apšviesti. Kita vertus, dviejų sluoksnių architektūroje naudojamas sudėtingesnis tinklas su papildomu sluoksniu, skirtu sudėtingesnėms vaizdų ypatybėms užfiksuoti. Dėl to, gilesnis tinklas gali demonstruoti aukštesnį tikslumą, ypač sudėtingesnėse situacijose (pvz., skirtingomis apšvietimo sąlygomis, susiliejusiuose kadruose, persidengimuose). Lyginant šiuos du metodus, siekiama rasti kompromisą tarp skaičiavimo efektyvumo ir aptikimo našumo.

Vieno sluoksnio architektūra – pasirinkta naudoti YOLOv7 dėl modelio gebėjimo apdoroti duomenis realiu laiku ir pateikiamų aukšto tikslumo rezultatų. Analizės metu pastebėta, kad modelis ypač populiarus panašaus tipo logotipų aptikimo ir atpažinimo uždaviniams spręsti, todėl nuspręsta modelį išbandyti su paruoštais duomenų rinkiniais. YOLOv7 modelis remiasi ELAN architektūra. ELAN geba sukurti efektyvų tinklą, kuriame valdo trumpiausią ir ilgiausią gradiento kelią, kad gilesni tinklai galėtų susiliesti ir efektyviai mokytis (žr. 33 pav.) Tam naudojami išplėtimo, maišymo ir sujungimo kardinalumai.



33 pav. YOLOv7 architektūros vizualizavimas [21]

YOLO modelio mokymui modifikuota keletą parametrų:

- įvesties vaizdo dydis;
- stuburo architektūra: pagrindinė architektūra apibrėžia bazinę tinklo architektūrą, naudojamą funkcijoms išgauti iš įvesties vaizdo. YOLOv7 galima naudoti populiarias pagrindines architektūras, tokias kaip CSPDarknet53 arba CSPResNeXt50;
- aptikimo sluoksnių skaičius: YOLOv7 turi kelis aptikimo sluoksnius, kurie numato skirtingų mastelių objektus. Aptikimo sluoksnių skaičius gali būti nurodytas atsižvelgiant į pageidaujimą tikslumo ir greičio kompromisą;
- filtrų skaičius: filtrų skaičių konvoliuciniuose sluoksniuose galima reguliuoti, kad būtų galima valdyti tinklo pajėgumą ir jo gebėjimą užfiksuoti sudėtingas funkcijas;
- pasitikėjimo slenkstis: pasitikėjimo slenkstis yra minimalus patikimumo balas, kurio reikia, kad objektas būtų laikomas tinkamu aptikimu. Jį galima reguliuoti, kad būtų galima valdyti tikslumo ir atšaukimo kompromisą;
- partijos dydis (angl. *Batch size*): mokymo pavyzdžių, perduodamų tinkle per vieną pirmyn ir atgal, skaičius. Jis nustato, kiek mėginių apdorojama lygiagrečiai prieš atnaujinant modelio svorius. Partijos dydžio pasirinkimas gali turėti įtakos mokymo trukmei, atminties reikalavimams ir modelio apibendrinimui. Didesni partijos dydžiai gali lemti greitesnį mokymą, bet reikalauja daugiau atminties, o mažesni partijų dydžiai gali užtikrinti geresnį apibendrinimą, bet lėtesnį mokymą;
- epochos: epocha reiškia visą treniruočių duomenų rinkinio perėjimą per modelį treniruotės metu. Kiekvienos epochos metu modelis pakartotinai apdoroja treniruočių duomenis paketais, apskaičiuoja nuostolius ir atnaujinama modelio svorius pagal pasirinktą optimizavimo algoritmą. Epochų skaičius lemia, kiek kartų modelis matys visą mokymo duomenų rinkinį. Per mažai epochų gali sukelti netinkamą pritaikymą, kai modelis nevisiškai užfiksuoja duomenų šablonus, o per daug epochų gali sukelti per daug pritaikymą, kai modelis tampa per daug specializuotas mokymo duomenims ir prastai veikia su naujais duomenimis.

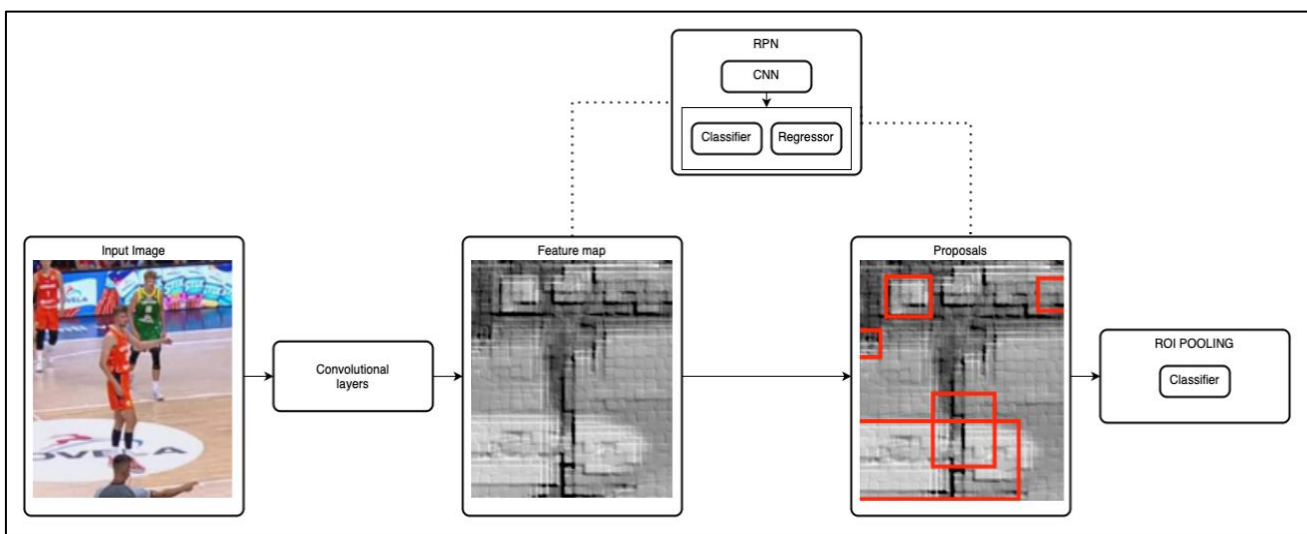
Dviejų sluoksnių architektūra – pasirinkta naudoti Greitesnį R-CNN dėl literatūros analizėje pastebėtų aukštų modelio tikslumo rezultatų aptinkant smulkius objektus. Taip pat modelis yra lankstus – galima pritaikyti įvairiems logotipo aptikimo scenarijams. Tai leidžia įtraukti įvairius pagrindinius tinklus, kurie suteikia galimybę subalansuoti modelio sudėtingumą ir našumą. Be to, jis palaiko kelių mastelių ir kelių objektų aptikimą, leidžiantį aptikti įvairaus dydžio logotipus ir kelis logotipus vaizde. Modelio architektūrą sudaro trys dalys (žr. **34 pav.**):

- konvoliuciniai sluoksniai – šiuose sluoksniuose mokome filtrus išgauti atitinkamas vaizdo savybes;
- regiono pasiūlymų tinklas (RPN) – mažas neuroninis tinklas paskutiniame konvoliucijos sluoksnių bruožų žemėlapyje, kuris nuspėja, ar yra objektas, ar ne;
- klasių ir ribinių langelių numatymo tinklas – kitas pilnai sujungtas neuroninis tinklas, kuris paima RPN siūlomus regionus ir prognozuoja objektų klasę.

Greitesnio R-CNN modelio mokymo metu galima modifikuoti keletą parametrų, tokių kaip:

- FNP architektūra: pagrindinis tinklas yra atsakingas už funkcijų ištraukimą iš įvesties vaizdo;
- inkaro dydžiai ir santykiai: greitesnis R-CNN naudoja iš anksto nustatytas skirtingo dydžio ir formato inkaro dėžes, kad pasiūlytų galimus objekto regionus;

- inkaro žingsnis: erdvinis žingsnis, kuriuo objektų žemėlapyje generuojami inkarai. Jis nustato inkaro dėžių erdvinę skiriamąją gebą ir daro įtaką objekto pasiūlymų dydžiui;
- IOU slenkstis: susikirtimo virš jungimo (IoU) slenkstis naudojamas teigiamiems ir neigiamiems inkaro dėžėms treniruotės metu nustatyti. Jis apibrėžia persidengimą, kurio reikia, kad inkaro dėžė būtų laikoma teigiamu arba neigiamu pavyzdžiu.
- klasifikavimo praradimas: klasifikavimo praradimo pasirinkimas, pvz., kryžminės entropijos praradimas arba židinio praradimas, naudojamas objekto klasifikatoriui mokytis;
- pradinis mokymosi greitis: pradinis mokymosi greitis, naudojamas mokant tinklą. Jis nustato žingsnio dydį optimizavimo metu;
- mokymosi greičio tvarkaraštis: mokymosi greičio koregavimo strategija treniruočių metu, pvz., žingsnio mažėjimas, eksponentinis mažėjimas arba ciklinis mokymosi greitis;
- optimizavimo funkcija: pasirenkamas optimizavimo algoritmas, pvz., stochastinis gradiento nusileidimas (SGD) arba Adam, naudojamas modelio svoriams atnaujinti.



**34 pav.** Greitesnio R-CNN architektūros vizualizavimas [22]

Tiriamosios dalies vertinimui pasitelktos metrikos:

- tikslumas (angl. *precision*) – modelio teigiamų prognozių tikslumo matas. Jis kiekybiškai įvertina teisingai numatytų teigiamų atvejų (tikrųjų teigiamų) proporciją iš visų atvejų, numatytų kaip teigiami (tikrieji teigiami + klaidingi teigiami). Kitaip tariant, tikslumas sutelkiamas į teigiamų prognozių teisingumą. Didelis tikslumas rodo, kad kai sistema numato teigiamą atvejį, tikėtina, kad jis bus teisingas, sumažinant klaidingų teigiamų atvejų skaičių. Mažas tikslumo balas rodo didesnę klaidingų teigiamų rezultatų tikimybę, o tai reiškia, kad sistema daro klaidingesnes teigiamas prognozes (2);
- atkūrimas (angl. *recall*) – taip pat žinomas kaip jautrumas arba tikras teigiamas rodiklis, matuoja modelio ar sistemos gebėjimą teisingai nustatyti teigiamus atvejus. Tai rodo teisingai numatytų teigiamų atvejų (tikrųjų teigiamų) proporciją iš visų faktinių teigiamų atvejų (tikrieji teigiami + klaidingai neigiami). Atkūrimas sutelkia dėmesį į kuo daugiau užfiksuotų teigiamų atvejų. Didelis atkūrimo balas rodo, kad sistema efektyviai nustato teigiamus atvejus ir sumažina klaidingų neigiamų atvejų skaičių. Žemas atkūrimo balas rodo, kad sistema nesugeba nustatyti daug faktinių teigiamų atvejų, todėl klaidingai neigiamų atvejų yra daugiau (3);

- F1 rezultatas – metrika, kuri sujungia tikslumą ir atkūrimą į vieną reikšmę, suteikianti subalansuotą dvejetainio klasifikavimo modelio našumo įvertinimą. Ji apskaičiuojama naudojant harmoninį tikslumo ir atšaukimo vidurkį. F1 balas paprastai naudojamas atliekant užduotis, kuriose klasių pasiskirstymas yra nesubalansuotas, todėl labai svarbu pasiekti tikslumo ir prisiminimo pusiausvyrą (4).
- mAP (angl. *mean Average Precision*) – matuoja modelio tikslumą ir atkūrimą lokalizuojant ir klasifikuojant objektus vaizde. Jis apskaičiuoja kiekvienos klasės vidutinį tikslumą (AP (5)) ir tada apskaičiuoja tų AP verčių vidurkį visose klasėse. AP yra tikslaus atkūrimo kreivės apibendrinimo metrika, kurioje tikslumas ir atkūrimas apskaičiuojami esant skirtingoms objekto aptikimo patikimumo slenksčiams. mAP pateikia bendrą modelio našumo įvertinimą keliose klasėse ir skirtinguose objektų aptikimo patikimumo lygiuose (6).

$$\text{Tikslumas} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Atkūrimas} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 * \frac{\text{tikslumas} * \text{atkūrimas}}{\text{tikslumas} + \text{atkūrimas}} \quad (4)$$

$$AP = \frac{1}{11} \sum_{\text{Atkūrimas}_i} \text{Tikslumas}(\text{Atkūrimas}_i) \quad (5)$$

$$mAP = \frac{\sum AP}{i} \quad (6)$$

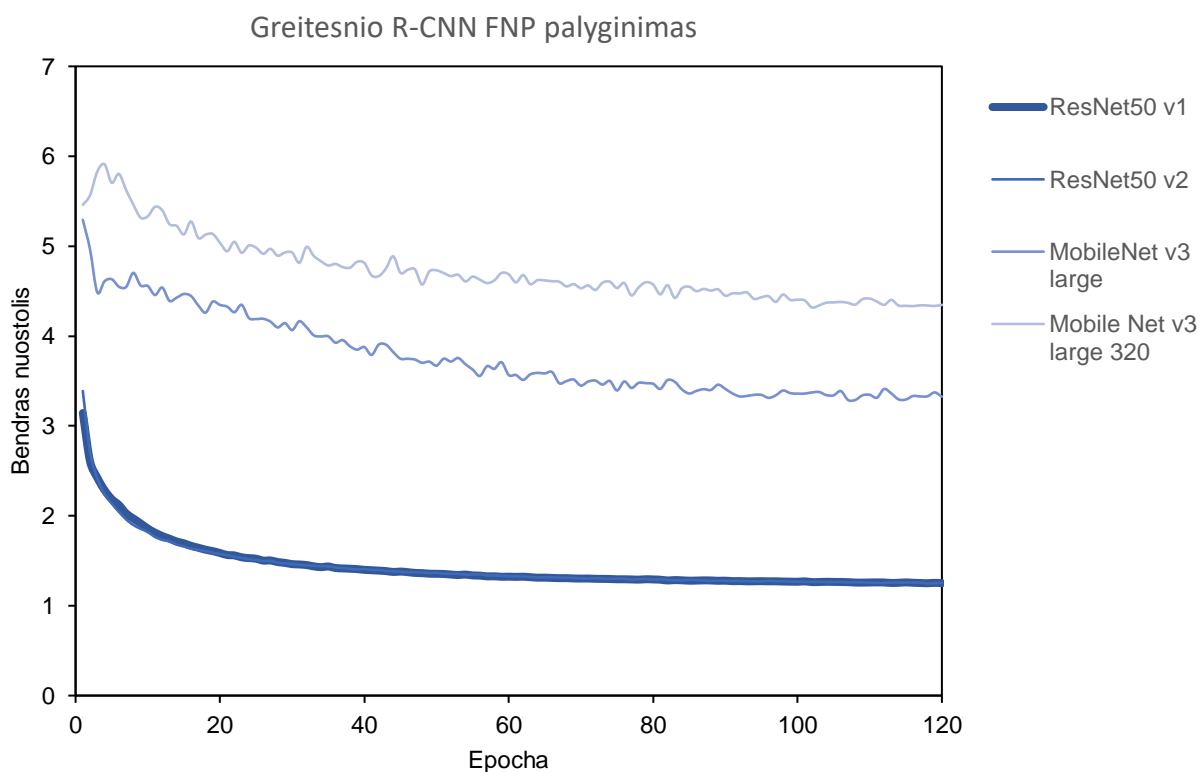
- klasifikatoriaus praradimas (angl. *Classifier Loss*): klasifikatoriaus praradimas yra susijęs su objekto detektoriaus klasifikavimo užduotimi – jis matuoja skirtumą tarp numatytų klasių tikimybių ir pagrindinės tiesos klasės etikečių vaizde esantiems objektams. Įprastos objektų aptikimo užduočių praradimo funkcijos apima *Softmax* ir *Negative Log Likelihood Loss*;
- dėžutės praradimas (angl. *Box Loss*): dėžutės praradimas yra susijęs su regresijos užduotimi, skirta ribojančių langelių prognozėms. Jis kiekybiškai įvertina skirtumą tarp numatytų ribojančių langelių koordinatų (pvz., x, y, plotis, aukštis) ir žemės tiesos ribojančiojo langelio koordinatų. Šiuo tikslu dažniausiai naudojamas sklandus L1 praradimas, nes jis yra mažiau jautrus nuokrypiams nei vidutinės kvadratinės klaidos (MSE) praradimas;
- RPN dėžutės praradimas (angl. *RPN BoxLoss*): regiono pasiūlymų tinklas (RPN) yra atsakingas už regionų pasiūlymų arba kandidatų objektų ribojamųjų langelių generavimą. RPN langelio praradimas matuoja skirtumą tarp numatytų ribojamojo langelio koordinatų iš RPN ir pagrindinio tiesos ribinio langelio koordinatų. Panašiai kaip dėžutės praradimas, dažnai naudojamas Smooth L1 loss.
- objekto praradimas (angl. *Object Loss*): objekto praradimas būdingas RPN ir matuoja, kaip gerai RPN atskiria priekinio plano (objekto) ir fono sritis vaizde. Šis praradimas padeda sugeneruoti tikslus regionų pasiūlymus objektams tolesniuose aptikimo vamzdyno etapuose. Objektyvumo prognozavimui dažniausiai naudojamas dvejetainis kryžminės entropijos praradimas;
- bendras nuostolis (angl. *Total Loss*): tai bendra nuostolių vertė, kurią modelis siekia sumažinti treniruočių metu. Paprastai tai yra kelių antrinių nuostolių derinys, įskaitant klasifikatoriaus

praradimą, langelio (ribojamojo langelio) praradimą, regiono pasiūlymo tinklo (RPN) langelį ir objektiškumo praradimą.

## 5.2. Logotipų aptikimo rezultatai

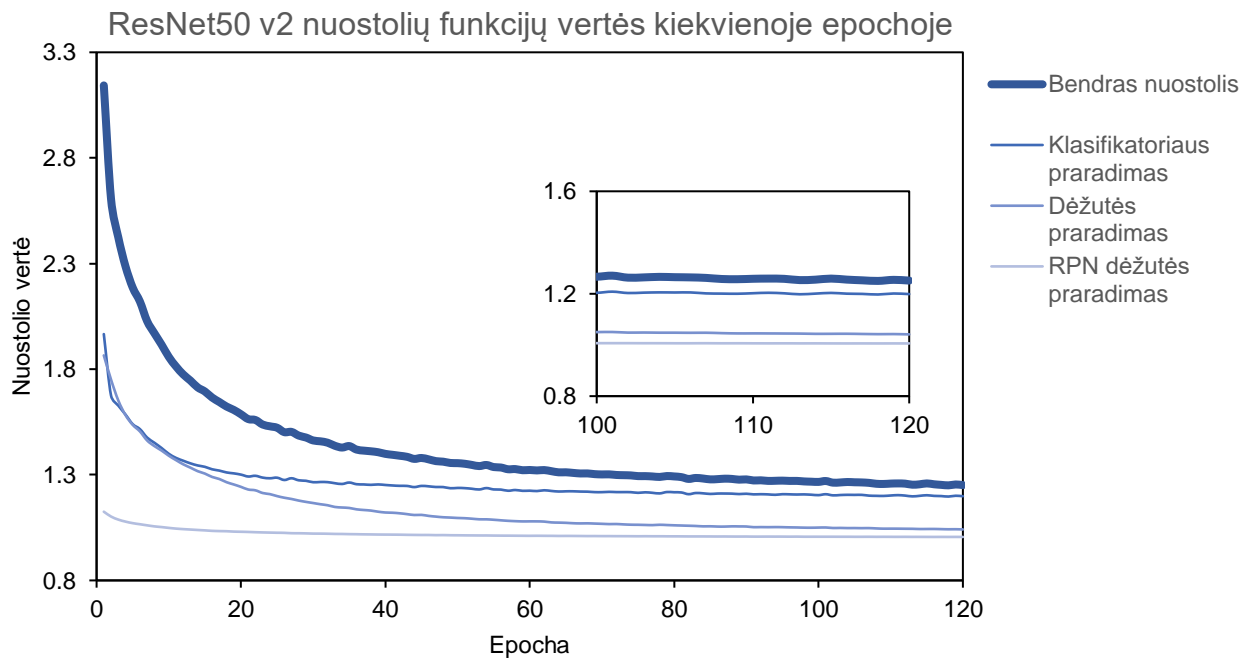
Tam, kad greitesnis R-CNN teiktų kuo tikslesnes objektų vietos prognozes, labai svarbu išsirinkti tinkamiausią funkcijų piramidės tinklą (FPN). Šio eksperimento tikslas – tarpusavyje palyginant populiarias objektų apkitimui naudojamas architektūras: *ResNet50*, *ResNet50 v2*, *MobileNet v3 Large* ir *MobileNet v3 large 320*.

Visi modeliai buvo mokyti 120 epochų. Pastebėta, kad *MobileNet* architektūros pateikė prasčiausius rezultatus, su didžiausiu bendru nuostoliu, o abi *ResNet* architektūros – labai panašų tikslumą (žr. **35 pav.**). Tačiau *ResNet50 v2*, dėl didesnio parametų skaičiaus, veikia kur kas lėčiau. Šie rezultatai pabrėžia, kad *ResNet50 v2* yra efektyviausias modelis šiai užduočiai, tačiau dėl jo greičio galima būtų svarstyti svarbą tarp tikslumo ir greičio.



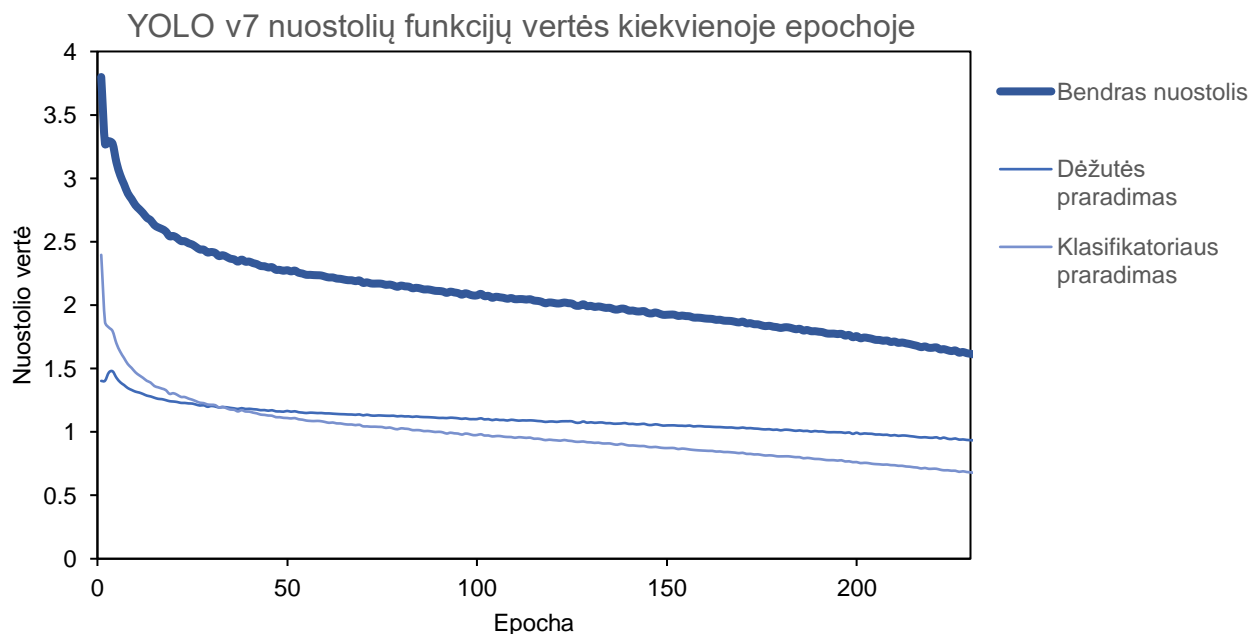
**35 pav.** Skirtingų FPN rezultatų palyginimas

Detaliau įvertinus greitesnio R-CNN modelio su *ResNet50 v2* FNP architektūra tikslumą pastebėta, kad mokymo proceso pradžioje bendras nuostolis siekia daugiau nei 2, tačiau sparčiai mažėja per pirmas 20 epochų, o modelio tikslumas nuo 70 epochos pradeda kisti labai lėta – tai rodo, kad modelis pradeda stabilizuotis. Iš modelio tikslumo kreivių (žr. **36 pav.**) galime pastebėti, kad didžiausią klaidos dalį sudaro modelio negebėjimas tiksliai įvardyti aptiktų objektų klasės, sekancioje pozicijoje – ne itin tikslus objektų dėžučių nustatymas.



36 pav. ResNet50 v2 nuostolių verčių kaita mokymo metu

Tuo pačiu principu įvertinami ir YOLO v7 gauti rezultatai. Svarbu paminėti, kad nuostolių reikšmių tarpusavyje lyginti negalime, dėl modeliuose taikomų skirtingų skaičiavimų paradigms. Mokymo proceso pradžioje bendras nuostolis siekia daugiau nei 3,5, tačiau sparčiai mažėja per pirmas 50 epochų, o modelio tikslumas nuo 200 epochos pradeda kisti lėta – modelis pradeda stabilizuotis (žr. 37 pav.). Skirtingai nei greitesniame R-CNN, modelio didžiausia nuostolio vertė gaunama dėl ne itin tikslaus objektų dėžučių nustatymo.



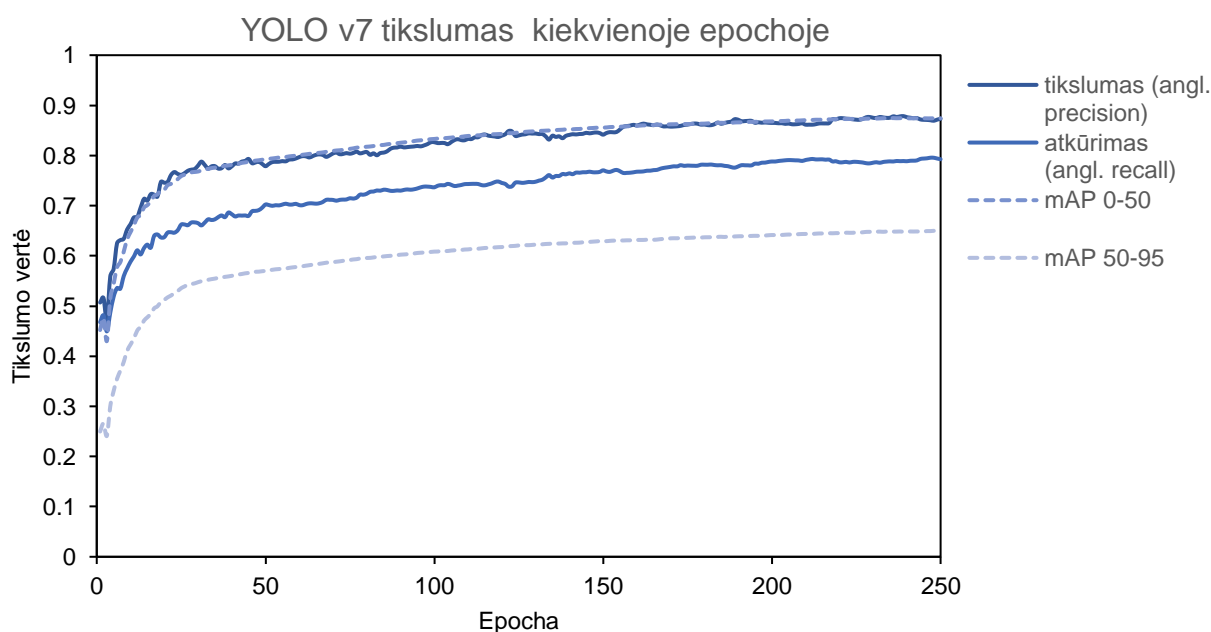
37 pav. YOLOv7 nuostolių verčių kaita mokymo metu

Detalesniam YOLO vertinimui pasitelkiamos tikslumo, atkūrimo bei mAP vertės (žr. 38 pav.). Per visą mokymo procesą šie rodikliai rodo reikšmingą pagerėjimą – tikslumas nuolat didėja ir pasiekia



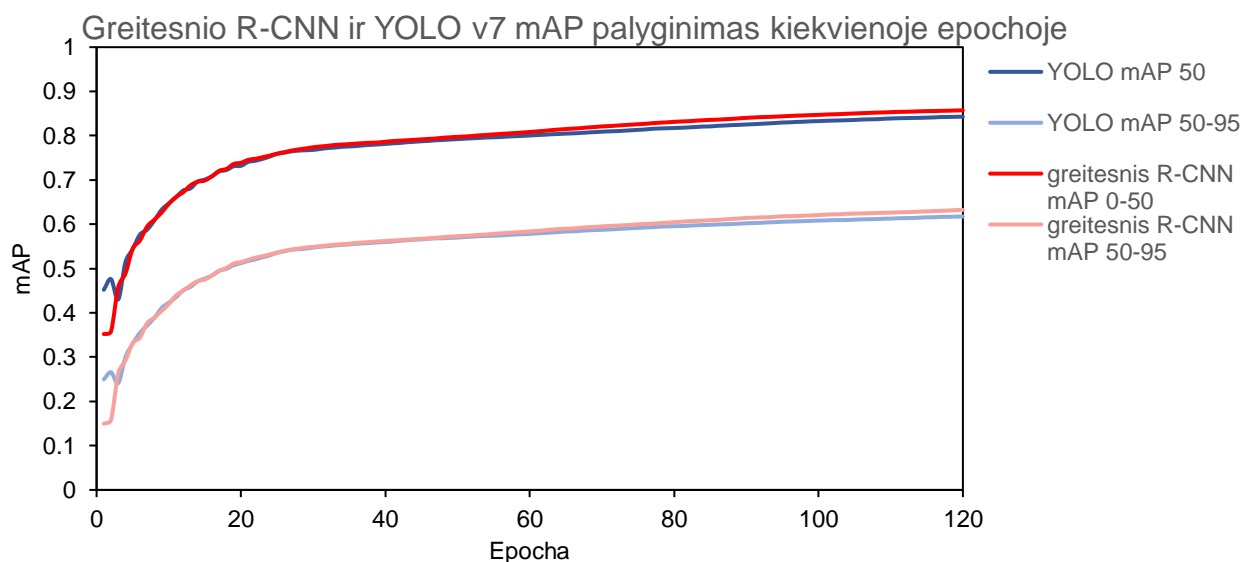
maždaug 0,9, o atkūrimas taip pat kyla ir stabilizuojasi ties 0,8. Tačiau nors abi metrikos laikui bėgant tobulėja, tikslumas nuolat viršija atkūrimą, o tai rodo, kad modelis pateikia mažiau klaidingų teigiamų objektų prognozių ir daugiau klaidingų neigiamų.

YOLO v7 modelis demonstruoja didelius mAP 0-50 ir mAP 50-95 tikslumo patobulinimus pradinio mokymo etapo metu. Pradžioje, tiek mAP 0-50, tiek mAP 50-95 rodikliai rodo staigų modelio tikslumo augimą. Iki apytiksliai 50 epochos, mAP 0-50 pasiekia apie 0.75, o mAP 50-95 - iki 0.6. Po šio taško modelio tikslumas ir toliau gerėja, tačiau daug lėčiau, kol galiausiai stabilizuojasi. Ši lėtesnė augimo fazė rodo, kad modelis artėja prie konvergencijos. MAP 0-50, rodantis tikslumą ties viena IOU riba, pasiekia didesnę galutinę vertę, stabilizuodamasis ties maždaug 0,85. Tai rodo, kad modelis yra labai tikslus identifikuojant teigiamus pavyzdžius su mažesniu tikslumo slenksčiu. Tuo tarpu, mAP 50-95, kuris nustato kelių slenksčių tikslumo vidurkį, stabilizuojasi ties maždaug 0,6. Šis rodiklis, nors ir žemesnis, pateikia išsamesnį vaizdą apie modelio gebėjimą tiksliai nustatyti objektus per įvairius slenksčius.



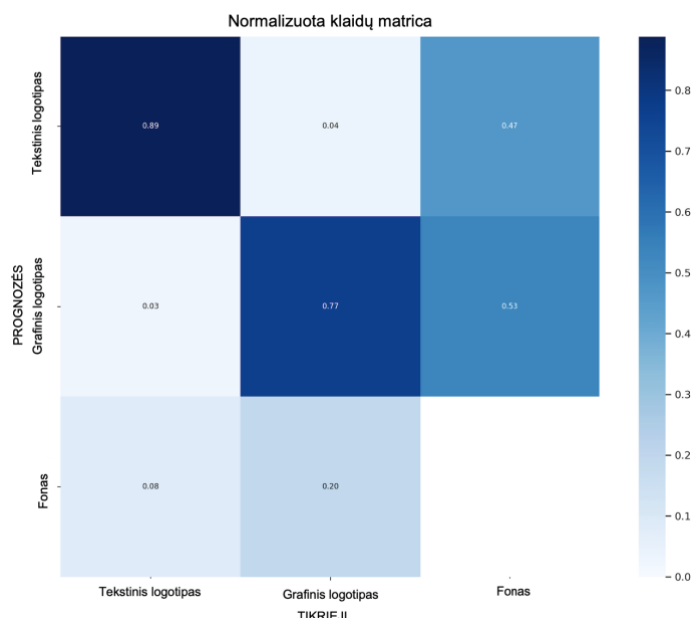
**38 pav.** YOLOv7 tikslumo verčių kaita mokymo metu

Greitesnis R-CNN ir YOLO modeliai lyginami naudojant vidutinio tikslumo (mAP) metriką. Įvertinę abiejų modelių mAP balus ir skirtingus IoU slenksčius, galime teigti, jog greitesnis R-CNN yra nežymiai pranašesnis – jo tikslumas lenkia 1,2% (YOLO 87,4%, greitesnis R-CNN 88,6%). Detalesnį skirtumą galime pastebėti paanalizavę grafiką (žr. **39 pav.**) (graikas pateikta tik iki 120 epochos dėl aiškesnio skirtumo tarp kreivių). Matome, kad nors greitesnis R-CNN šiek tiek atsilieka pradinėse mokymo epochose, vis dėlto pralenkia YOLO modelį. Greitesnis R-CNN paprastai užtikrina didesnę tikslumą dėl dviejų pakopų aptikimo struktūros, kur pirmiausia pasiūlomi regionai, o paskui jie klasifikuojami. Tuo tarpu YOLO yra tik vienos pakopos detektorius, kuris tiesiogiai nuspėja klasių tikimybes ir ribojančius langelius per vieną sluoksnį. Vertinant modelius tarpusavyje svarbi ir greitaveika – čia YOLO modelis rodo ryškų pranašumą. Atliktame tyrime, kuriame abu modeliai turėjo pateikti prognozes 10 000 paveikslėlių, YOLO modelis pademonstravo ženklų pranašumą – rezultatus pateikė net 13% greičiau.

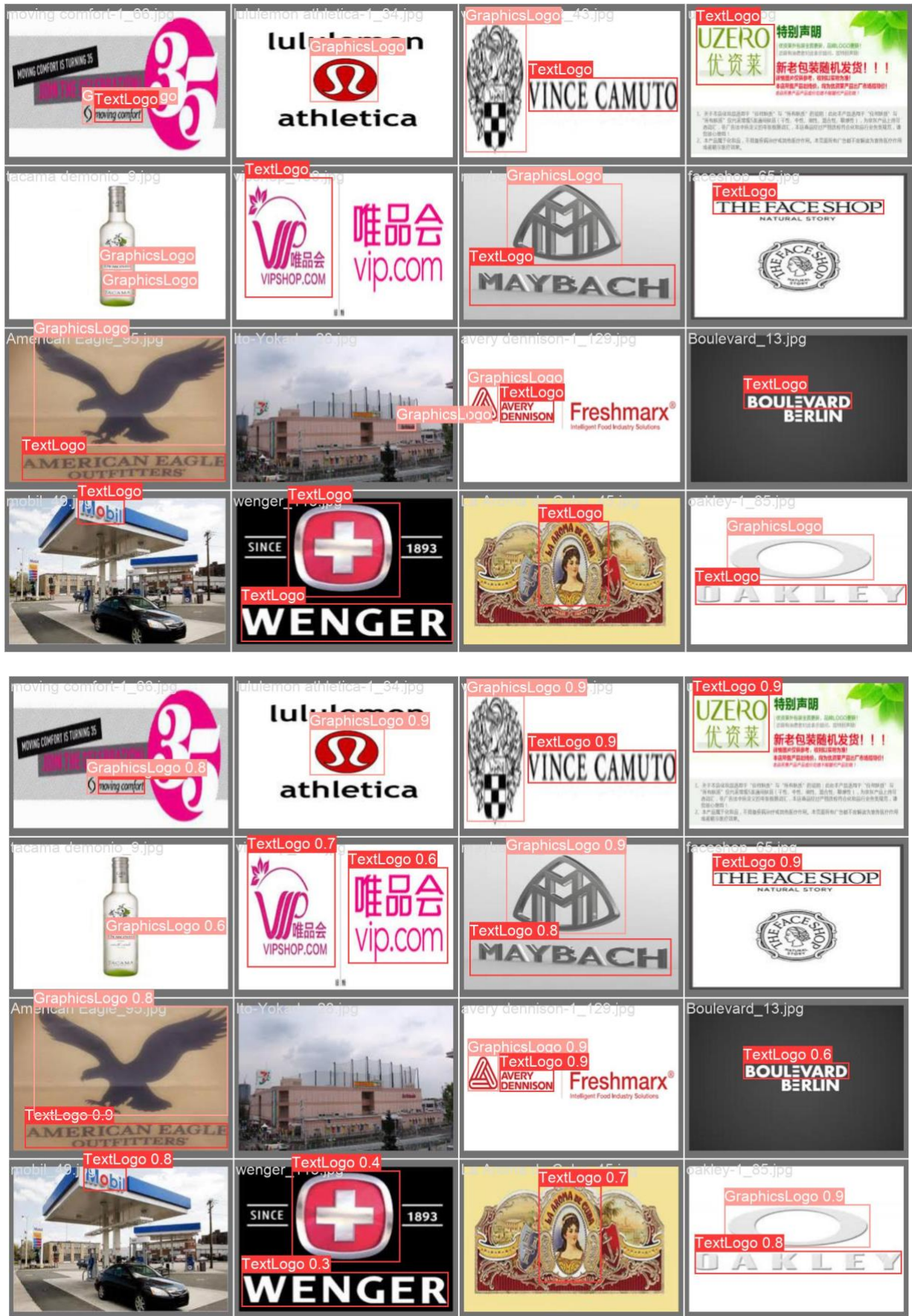


39 pav. Greitesnio R-CNN ir YOLOv7 tikslumo verčių palyginimas

Analizuojant modelio gebėjimą teisingai priskirti aptiktiems objektams klases, pasitelkiama klaidų matrica (žr. 40 pav.). Tai labai populiari mašininio mokymosi statistikos atvaizdavimo priemonė, kurios pagrindinis tikslas yra pateikti išsamų modelio kategorijų priskyrimo suskirstymą. Klaidų matrica atskleidžia, kad modelis geriausiai identifikuoja „Tekstinis logotipas“ – 89 % tikslumu. Tačiau galime identifikuoti ir modelio silpnąją savybę – atskirti grafinius logotipus ir foną (tai rodo žemesni teisingi klasifikavimo rodikliai – atitinkamai 77% ir 53%). Taip pat didelė dalis tekstinių logotipų yra klaidingai klasifikuojami kaip fonas (47 %), o nemaža dalis fono pavyzdžių klaidingai priskiriami grafiniams logotipams (20 %). Realius aptikimo scenarijus iliustruoja pateikti aptikimo pavyzdžiai (žr. 41 pav.).



40 pav. YOLOv7 klaidų matrica



41 pav. Logotipų duomenų rinkinio anotacijos (viršuje) bei prognozės (apačioje)

Šioje dalyje vertinamas modelių gebėjimas aptikti logotipus atskiruose atviruose duomenų rinkiniuose pritaikant tam tikras paveikslėlių modifikacijas. Gautų modelių mokymo įvestys aprašytos lentelėse.

YOLOv7 rezultatai su FlockLogos32 duomenų imtimi. Rezultatuose galime pastebėti, kad aukščiausią tikslumą pavyko pasiekti naudojant spalvų erdvės verčių normalizavimą, tačiau bendras tikslumas šiuo metodu pats žemiausias. Aukščiausią bendrą tikslumą pateikė nekeisti paveikslėliai (žr. **10 lentelė**).

**10 lentelė.** YOLOv7 rezultatai su FlickrLogos-32

Sąlygos	Tikslumas	Atkūrimas	F1	mAP
<b>Originalūs paveikslėliai</b>	<b>0.72</b>	<b>0.75</b>	<b>0.73</b>	<b>0.76</b>
512 x 512	0.65	0.6	0.62	0.68
Spalvų erdvės verčių normalizavimas	0.73	0.62	0.68	0.67
Spalvų erdvės konvertavimas į HSV	0.63	0.58	0.64	0.69

Greitesnio R-CNN rezultatai su FlockLogos32 duomenų imtimi. Rezultatuose galime pastebėti, kad aukščiausią tikslumą ir bendrą tikslumą pavyko pasiekti naudojant nepadorotus originalius paveikslėlius. Prasčiausi rezultatai gauti apkarpančiais paveikslėlius iki 512 x 512 dydžio (žr. **11 lentelė**).

**11 lentelė.** Greitesnio R-CNN rezultatai su FlickrLogos-32

Sąlygos	Tikslumas	Atkūrimas	F1	mAP
<b>Originalūs paveikslėliai</b>	<b>0.79</b>	<b>0.72</b>	<b>0.75</b>	<b>0.77</b>
512 x 512	0.59	0.70	0.64	0.66
Spalvų erdvės verčių normalizavimas	0.65	0.58	0.57	0.69
Spalvų erdvės konvertavimas į HSV	0.75	0.78	0.77	0.76

YOLOv7 rezultatai su *LogoDet-3K* duomenų imtimi. Rezultatuose galime pastebėti, kad nei viena modifikacija negerina rezultatų, gautų su neapdorotais paveikslėliais. (žr. **12 lentelė**).

**12 lentelė.** YOLOv7 rezultatai su *LogoDet-3K*

Sąlygos	Tikslumas	Atkūrimas	F1	mAP
<b>Originalūs paveikslėliai</b>	<b>0.76</b>	<b>0.73</b>	<b>0.74</b>	<b>0.75</b>
512 x 512	0.66	0.71	0.70	0.71
Spalvų erdvės verčių normalizavimas	0.61	0.49	0.54	0.58
Spalvų erdvės konvertavimas į HSV	0.62	0.69	0.61	0.64

Greitesnio R-CNN rezultatai su *LogoDet-3K* duomenų imtimi. Rezultatuose pastebime tą pačią tendenciją kaip ir YOLO rezultatuose. Objektai originaliuose paveikslėliuose aptinkami tiksliausiai. (žr. **13 lentelė**).

**13 lentelė.** Greitesnio R-CNN rezultatai su *LogoDet-3K*

Sąlygos	Tikslumas	Atkūrimas	F1	mAP
<b>Originalūs paveikslėliai</b>	<b>0.75</b>	<b>0.78</b>	<b>0.77</b>	<b>0.77</b>
512 x 512	0.64	0.67	0.65	0.66

Spalvų erdvės verčių normalizavimas	0.69	0.68	0.69	0.67
Spalvų erdvės konvertavimas į HSV	0.71	0.69	0.7	0.7

Pastebėta, kad modelio tikslumas nėra pakankamai aukštas dėl duomenų rinkiniuose esančių sunkiai atpažįstamų prekių ženklų. Pavyzdyje (žr. **42 pav.**) matome, kad prekės ženklas „Nike“ nėra aptinkamas dėl pernelyg didelio išsiliejimo į aplinką arba yra klaidingai atpažįstamas panašios formos klaidingi objektai. Tokių duomenų rinkiniuose yra labai daug. Pagrindinės priežastys:

- Prekių ženklai neišsiskiria iš aplinkos – jie išsilieję į ją
- Prekių ženklai sumaišomi su panašios formos, spalvos kitais objektais
- Prekių ženklai persidengia su kitais objektais
- Prekių ženklai dėl apšvietimo pakeičia spalvą, yra sunkiai matomi



**42 pav.** Keblūs aptikimo atvejai

Modelių tikslumo gerinimui pasitelktos tam tikros transformacijos. Papildymo seka apima keletą operacijų, strategiškai parinktų siekiant padidinti duomenų rinkinio įvairovę ir patikimumą. Atsitiktinio dydžio dėžės apkarpymas, nustatytas su 0,4 tikimybe, užima svarbų vaidmenį išlaikant logotipų ribojančių dėžių vientisumą atsitiktinai apkarplant ir keičiant dydį. Ši operacija yra būtina norint išsaugoti tikslią erdvinę informaciją ir užtikrinti, kad modelis išmoktų aptikti logotipus įvairiose dydžio skalėse. Be to, horizontalaus ir vertikalios apvertimo operacijos, kurių kiekvienos tikimybė 0,5, taikomos atskirai, kad būtų pajvairintas požiūrio taškas. Ši transformacija padeda modeliui nagrinėti skirtingas perspektyvas, pagerindama jo gebėjimą efektyviau apibendrinti duomenis. Į padidinimo strategiją taip pat įtrauktos atsitiktinės ryškumo ir kontrasto operacijos, kurių tikimybė yra 0,2, taip prisidedant prie įvairių apšvietimo sąlygų ir spalvų variacijų duomenų rinkinyje. Šiomis transformacijomis siekiama atskleisti modelį įvairiems scenarijams, pagerinant jo pritaikomumą ir tvirtumą.

Šių transformacijų dėka, Greitesnio R-CNN modelio tikslumas pagerėjo 1.1%, o YOLOv7 – 0.7%.

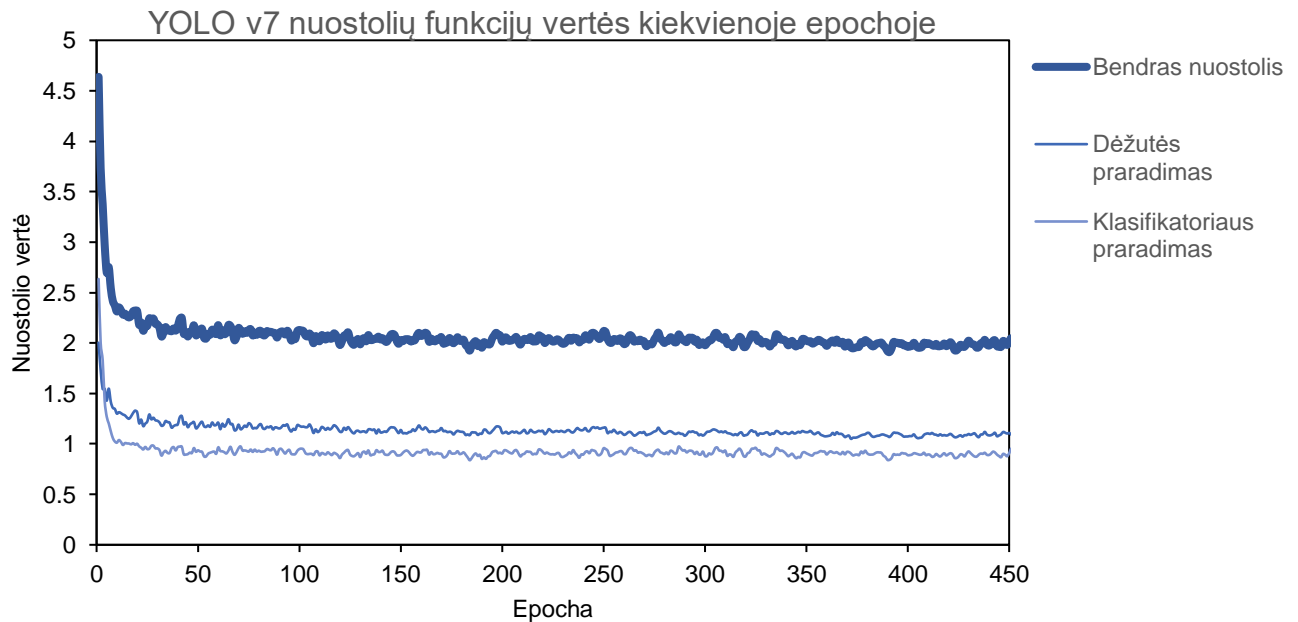
### 5.3. Prekių ženklų reklamų aptikimo tikslumo rezultatai krepšinio rungtynių kadruose

Sukurti modeliai ištestuoti naudojant sukurtą krepšinio rungtynėse matomų prekių ženklų reklamų anotacijų duomenų rinkinį. Nors logotipų rinkiniuose modeliai demonstravo sąlyginai aukštą tikslumą (mAP 0-50 0,85, mAP 50-100 0,6), vis dėlto šiame duomenų rinkinyje tikslumo rezultatai nėra aukšti – modelio mAP 0-50 siekia tik 0,42. Tuo galime įsitikinti patektame pavyzdyje (žr. 43 pav.) – nors tam tikri prekės ženklai aptikti sėkmingai, didelė dalis jų arba aptikti klaidingai, arba visai neaptikti.



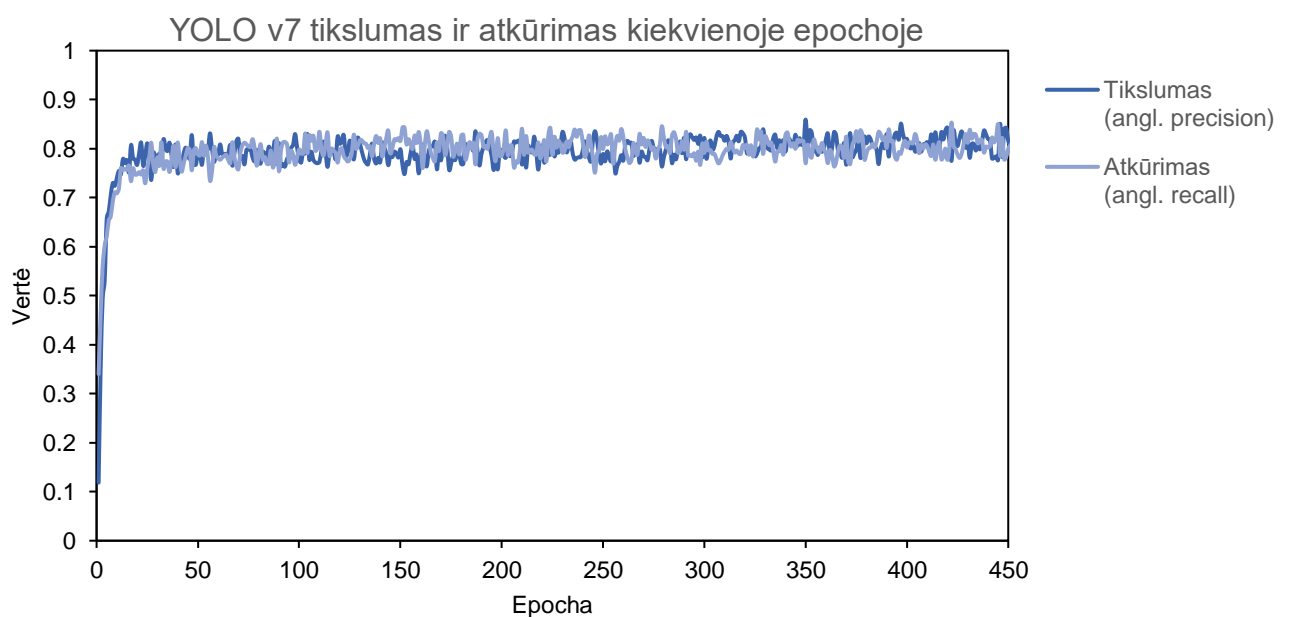
43 pav. Krepšinio rungtynių anotacijos (viršuje) bei prognozės (apačioje)

Rezultatų gerinimui esami modeliai papildomai mokomi panaudojant sukurta krepšinio rungtynėse matomų prekių ženklų reklamų anotacijų duomenų rinkinį. Modelis buvo mokomas 400 epochų, tačiau po 200 epochų modelis konvergavo ir tikslumas gerėjo itin nežymiai. Rezultatų tendencijos išlieka panašios – modelis mažiau klysta priskirdamas objektui klasę nei tiksliai aptikdamas jo padėtį (žr. **44 pav.**).



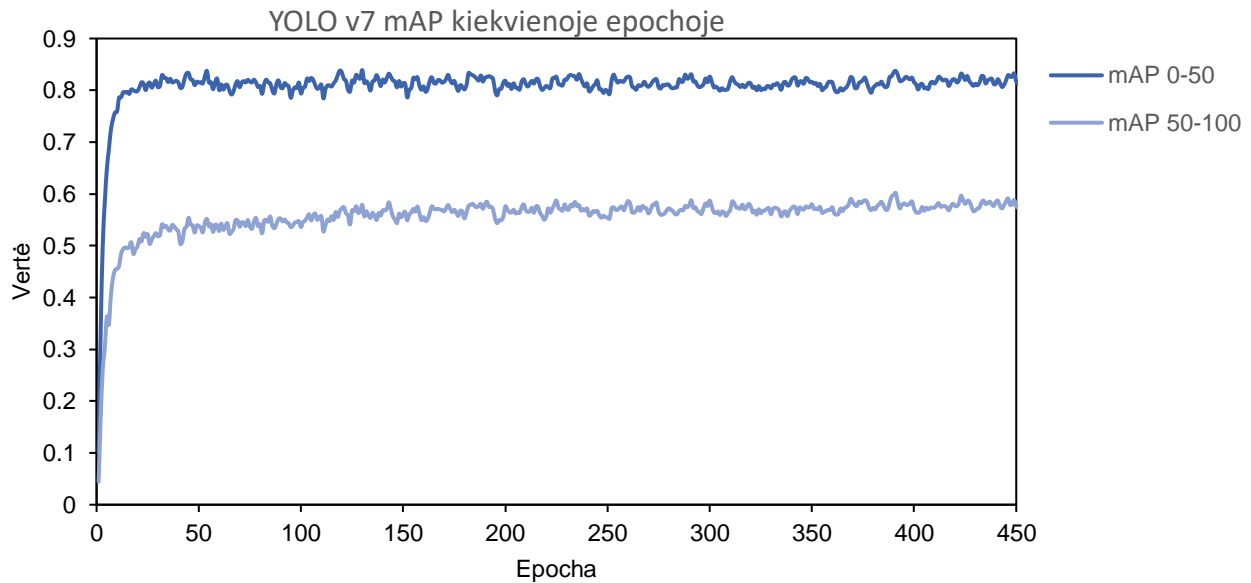
**44 pav.** YOLOv7 nuostoliai mokymo metu

Skirtingai nei modelio mokyme su logotipų duomenimis, kur tikslumas mokymo metu nuolat viršijo atkūrimą, pastebime labai panašius tikslumo ir atkūrimo rezultatus viso mokymo metu (žr. **45 pav.**). Tai reiškia, kad modelis vienodai gerai priskiria teigiamas klasifikacijas ir teisingai klasifikuoja teigiamus atvejus.



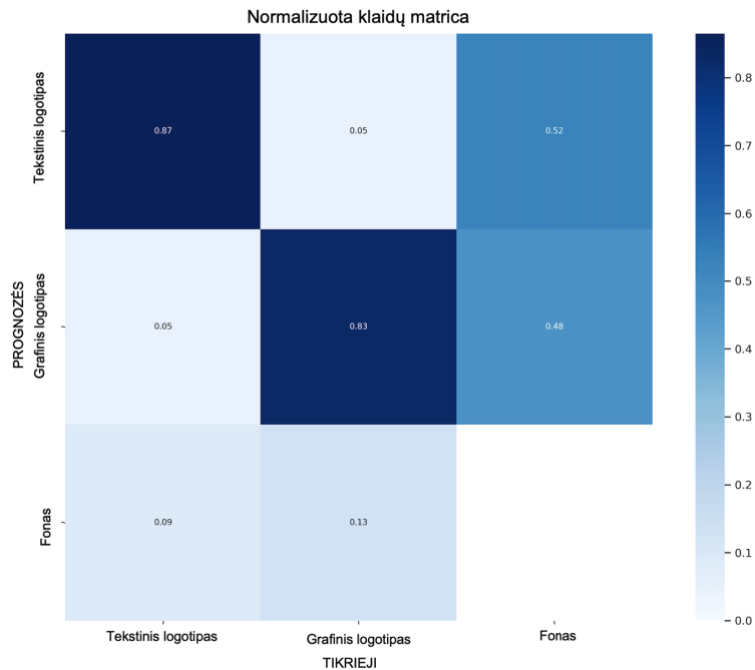
**45 pav.** YOLOv7 tikslumo verčių kaita mokymo metu

MAP 0-50 kreivė, kuri išlieka nuolat aukštesnė, stabilizuojasi ties apytiksliai 0,8 po 100 epochų, o tai rodo, kad modelis geba aptikti ir klasifikuoti objektus esant žemesniam IoU slenksčiui pakankamai tiksliai. Priešingai, mAP 50-95 kreivė viso mokymo metu nuolat palengva auga iki 0,55, o tai rodo, kad pasiekti aukštą tikslumą, esant aukštesnėms IoU slenksčiams, yra sudėtingiau (žr. **46 pav.**).



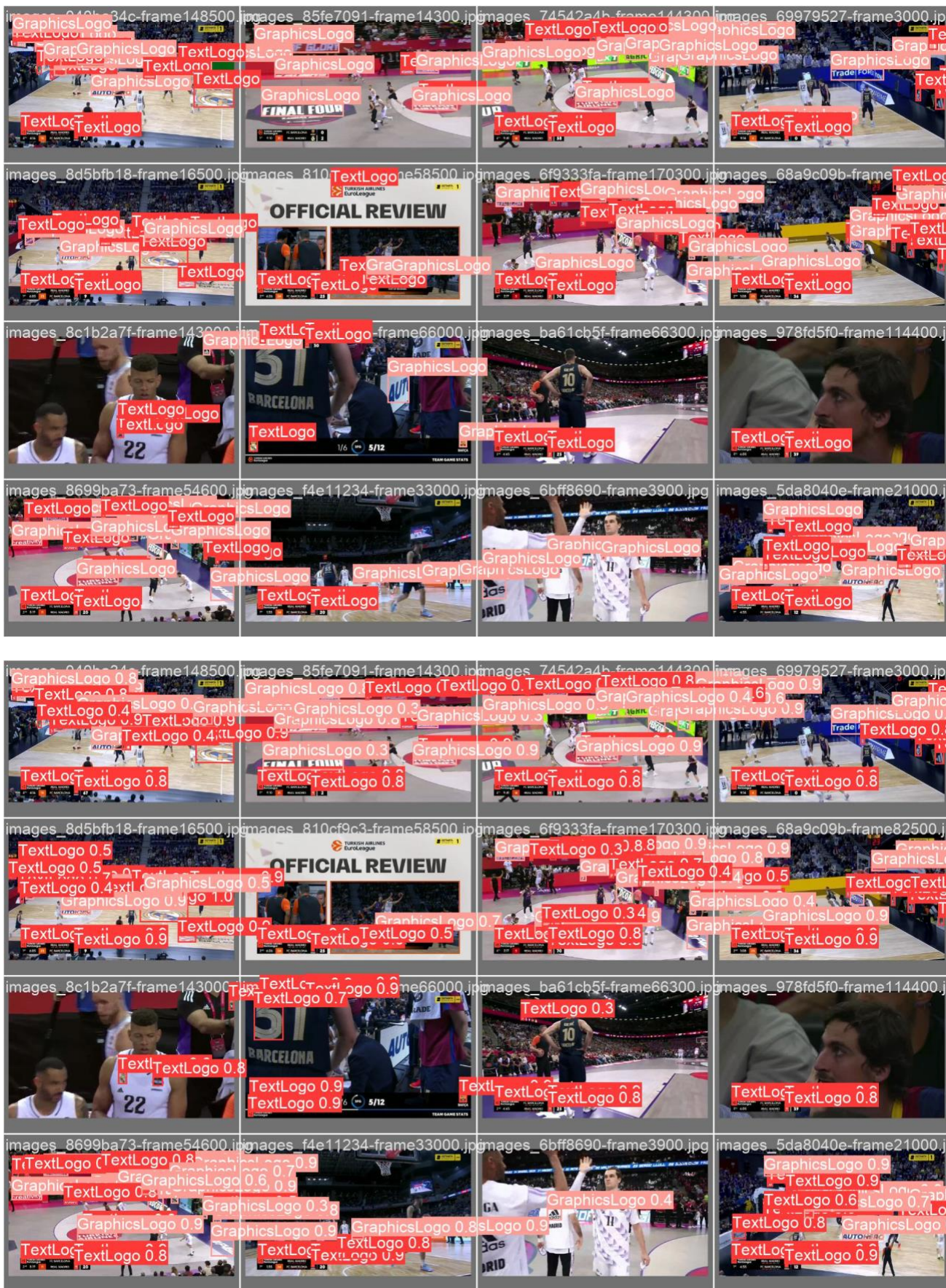
**46 pav.** YOLO bendro tikslumo su skirtingais IOU slenksčiais reikšmių kaita mokymo metu

Klaidų matricos rezultatai rodo (žr. **47 pav.**), kad modelis geba identifikuoti tiek tekstinius, tiek grafinius logotipus aukštu tikslumu, atitinkamai 87% ir 83%. Vis dėlto, didelė dalis fono objektų išlieka klaidingai identifikuojami kaip prekės ženklų reklamos – modelis linkęs teikti klaidingus teigiamus rezultatus. Realius aptikimo rezultatus iliustruoja pateiktas aptikimo pavyzdys (žr. **48 pav.**).



**47 pav.** YOLOv7 klaidų matrica



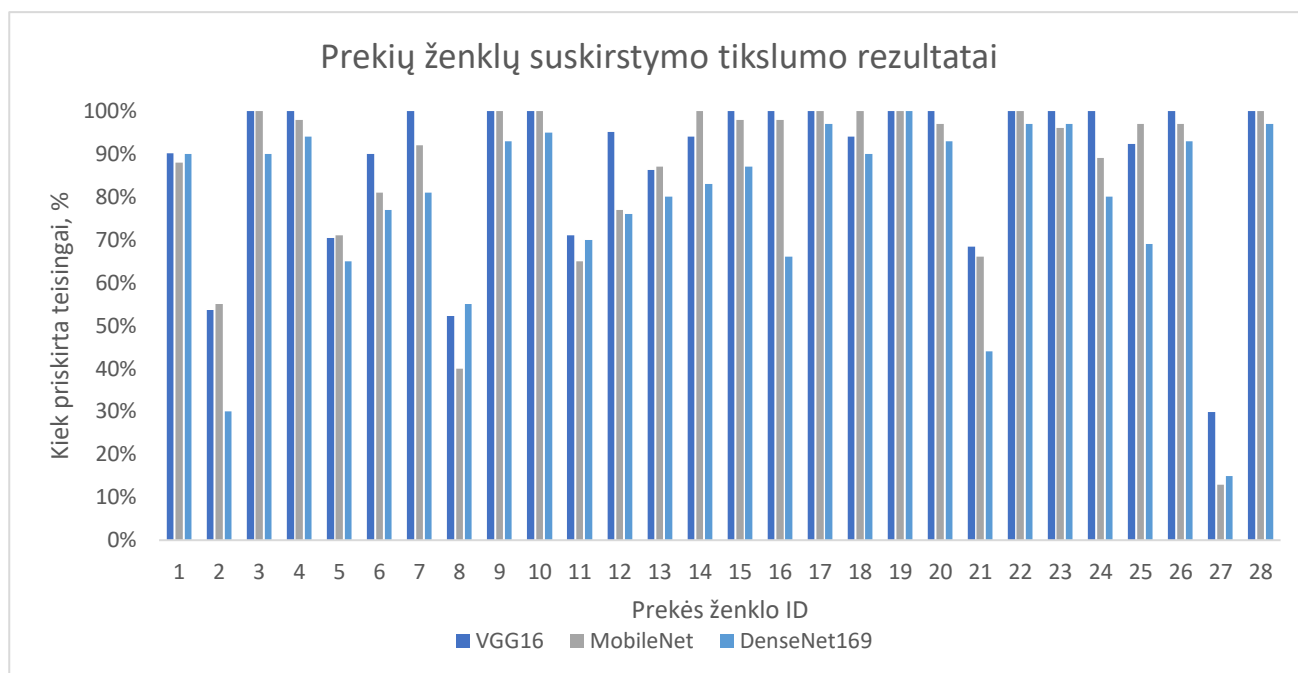


48 pav. Krepšinio rungtynių anotacijos (viršuje) bei aptikimo rezultatai po papildomo mokymo (apačioje)

#### 5.4. Prekių ženklų atpažinimo bei klasifikavimo testavimo rezultatai

Šiuo momentu, modelis geba tik aptikti prekių ženklų reklamų vietą bei priskirti jai klasę – ar tai tekstinis prekės ženklas, ar grafinis. Prekių ženklų reklamos efektyvumo vertinime svarbu vertinti kiekvieną prekių ženklą atskirai, todėl tolimesniame etape vertinamas modelio gebėjimas suskirstyti prekių ženklus pagal vaizdines savybes (žr. 2.4 Sistemos veikimo detalizavimas).

Eksperimento metu buvo panaudota 10 000 krepšinio rungtynėse aptiktų prekių ženklų. Jie buvo suskirstyti į 30 skirtingų klasterių. Testavimo metu pasitelktas rankinis vertinimas – kiekviename klasteryje suskaičiuoti neteisingai priskirti prekės ženklai. Vaizdinėms savybėms išgauti pasitelktos trys skirtingos populiaros architektūros – VGG16, MobileNet bei DenseNet169. Gauti rezultatai pateikti grafike (žr. 49 pav.). Pastebėta, kad dviejuose klasteriuose buvo sugrupuoti aptikti ne prekiniai ženklai, o fonas. Šie klasteriai į rezultatus neįtraukti. Galime pastebėti, kad daugeliu atvejų gauti tikslumo rezultatai siekia 80 % ar daugiau. Vis dėlto, tam tikrų prekės ženklų modelis teisingai sugrupuoti negeba dėl persidengimo, judesio ar pasikeitusios projekcijos. Taip pat galime pastebėti, kad VGG16 lenkia kitas architektūras ir pateikia tikslesnius rezultatus. Bendras tikslumas gautas su kiekviena architektūra pateiktas lentelėje (žr. 14 lentelė.).



49 pav. Prekių ženklų suskirstymo pagal vaizdines savybes eksperimento rezultatai

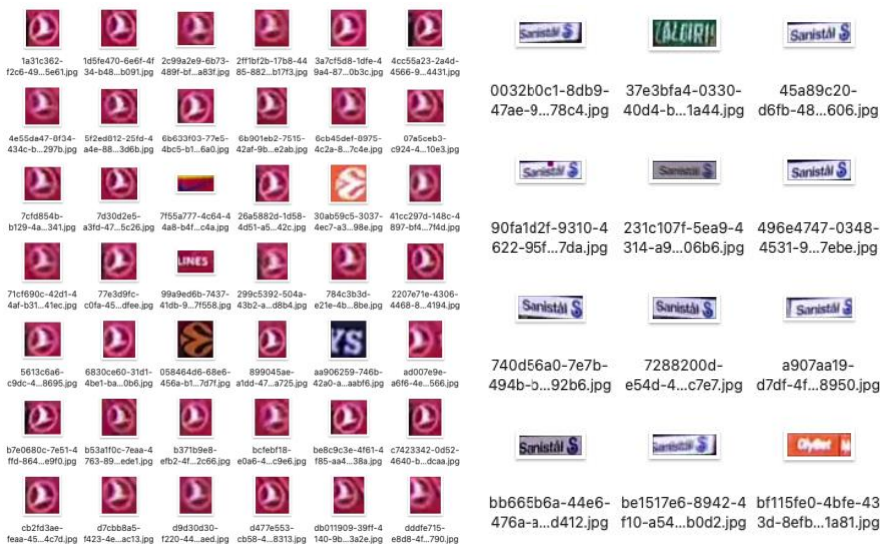
14 lentelė. Prekių ženklų grupavimui taikytų vaizdo savybių išgavimo architektūrų bendro tikslumo rezultatai

Vaizdinių savybių išgavimo modelis	Bendras tikslumas
VGG16	89 %
MobileNet	86 %
DenseNet169	79 %

Detaliau galime apžvelgti kaip atrodo skirtingi tikslumo prekių ženklų grupės (žr. 50 pav.). Matome, kad tam tikrus prekės ženklus modelis geba priskirti tam pačiam prekės ženklui net ir esant fono pasikeitimui ar persidengimui, tačiau tam tikrų prekės ženklų savybės gali sutapti ir jų pasiskirstymas tampa klaidingas.



### Labai geras suskirstymas



### Vidutinis suskirstymas



### Prastas suskirstymas

50 pav. Prekių ženklų suskirstymo rezultatų vizualizavimas

## 6. Prekių ženklų reklamos efektyvumo vertinimo sistemos naudotojo vadovas

Modelio pritaikymas naudojant *Python* programavimo kalbą pateikiamas faile *test\_model.py*. Modelio įvestys:

- `model_path` – modelio failas kompiuteryje;
- `video_path` - vaizdo įrašo failas kompiuteryje.
- `result` - rezultatų saugojimo vieta kompiuteryje.

Modelio rezultatai saugomi CSV formatu (žr. **51 pav.**):

- prekės ženklo id;
- prekės ženklo matomumo trukmė;
- prekės ženklo užimamas plotas;
- prekės ženklo anamorfinė projekcija;
- prekės ženklo ploto dalies santykis su kitų prekės ženklų plotais;
- Prekės ženklo nutolimas nuo centro.

Results

ID	Duration	Covered area	Projection index	Area coverage ratio
60fd8400-de40-417f-ab99-1cf338e752df	209	9238	0.32	0.1
c6942185-941c-468f-9091-1577592d0875	358	6932	0.83	0.08
330e4b21-103a-4462-b2ad-22a3e66ef831	241	5908	0.54	0.04
b6fb0f33-6a91-40c5-b901-aa0bec1c1b51	395	7175	0.19	0.04
0dfb558c-3629-4078-b1c9-50db03eecd56	539	4897	0.89	0.04
9f606987-76b9-4895-82f8-a9b70c450a82	325	3506	0.92	0.01
67440686-9bf5-4adb-ad71-754cb3c2904f	424	6266	0.18	0.07
d5581dbb-d72a-42d9-a4e1-f2d8ffad2091	351	5415	0.45	0.06
72cf5326-1baa-48d7-8089-420ef10eb567	467	7264	0.26	0.09
c3090bd8-2fc5-4237-9950-5d6fb2b4d30a	458	7729	0.54	0.08

**51 pav.** Rezultatų failo pavyzdys

Pradėjus algoritmą, pasirinktas vaizdo įrašas yra analizuojamas, jos metu galima analizės proceso peržiūra. Vaizdo įrašo analizės metu prekių ženklai yra sekami, jų judėjimo pozicijos vaizduojamos kreivėmis. Kiekvienai aptiktai prekės ženklo reklamai suteikiamas unikalus ID (žr. **52 pav.**).



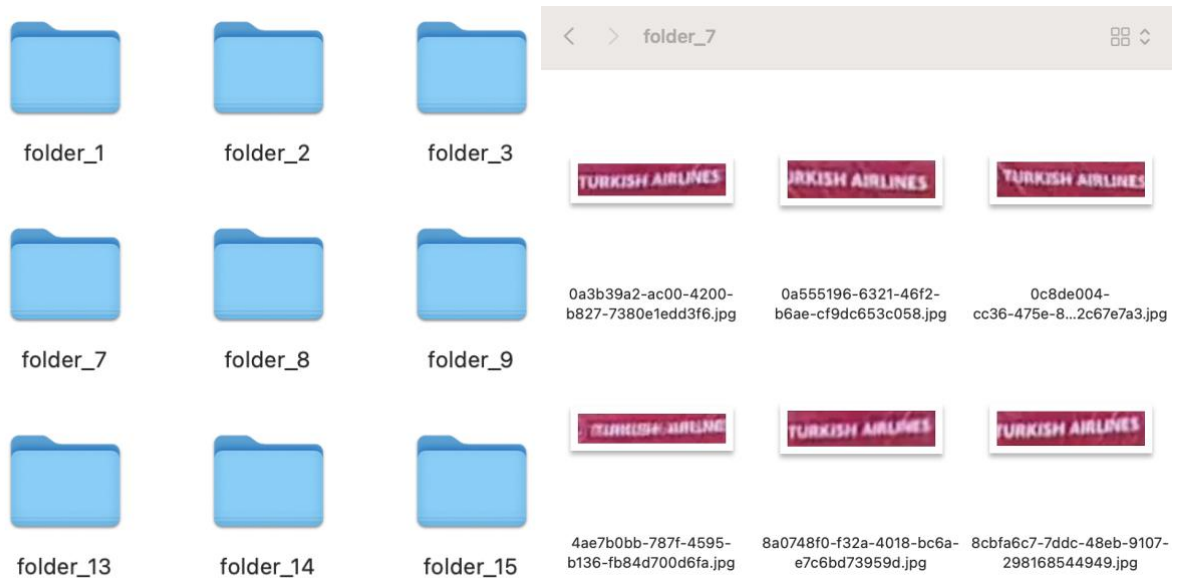
52 pav. Prekių ženklų reklamų aptikimas ir sekimas

Modelio aptikti prekės ženklai saugomi pasirinktame aplanke. Kiekvienam aptiktam prekės ženklui suteikiamas unikalus ID (žr. 53 pav.).



53 pav. Aptikti prekės ženklai išsaugoti aplanke

Gauti modelio aptikimo rezultatai apdorojami faile *process\_data.py*. Visų pirma, aptikti prekės ženklai yra suskirstomi į grupes pagal vaizdines savybes. Kiekviena grupė yra laikoma tuo pačiu prekės ženklu (žr. 54 pav.). Duomenų faile, kuriame saugomos visų aptiktų prekių ženklų matomumo metrikos, atrenkami kiekvieno prekės ženklo grupės paveikslėliai pagal jų ID. Galiausiai, pateikiami apdoroti rezultatai.



**54 pav.** Prekių ženklų grupės ir grupės rezultatai

## Išvados

1. Atlikus analizę nustatyta, kad *LogoDet-3K* duomenų rinkinys yra pats tinkamiausias uždavinio sprendimui, nes jį sudaro įvairus spektras skirtingų prekių ženklų su skirtingomis jų variacijomis. Atlikta pasirinkto duomenų rinkinio analizė po duomenų rinkinio apdorojimo leido nustatyti, kad prekių ženklų, kurie klasifikuojami kaip tekstiniai, yra daugiau nei grafinių, todėl prieš modelio apmokymą prireikia atlikti duomenų rinkinio balansavimą.
2. Tyrimo metu, kuriuo siekta išsiaiškinti, kuri greitesnio R-CNN FNP teikia aukščiausius tikslumo rezultatus, pastebėta, kad tiek *ResNet50*, tiek *ResNet50 v2* pateikia panašius aukščiausius tikslumo rezultatus, kitas architektūras lenkdamas net 10 kartų.
3. Lyginant YOLOv7 ir Greitesnio R-CNN gautus rezultatus pastebėta, kad greitesnis R-CNN pateikia 1,2 % tikslesnius aptikimo rezultatus, tačiau veikia net 13 % lėčiau. R-CNN tikslumo pranašumas pastebimas tikslesniuose smulkių prekių ženklų aptikimo rezultatuose. Nustatyta, kad visų kadrų dydžio keitimą į 512 x 512 ar spalvų erdvės konvertavimas modelio tikslumo negerina. Atlikus atsitiktines paveikslėlių apkarpymo, padėties keitimo bei ryškumo transformacijas pastebėtas 1,1 % greitesnio R-CNN bei 0,7 % YOLOv7 tikslumo pagerėjimas.
4. Dėl itin lėto greitesnio R-CNN apdorojimo, tolimesnius tyrimus nuspręsta tęsti tik su YOLOv7. Juose nustatyta, kad modelis, mokytas naudojant tik atvirus logotipų duomenų rinkinius, geba aptikti prekės ženklus krepšinio rungtynių kadruose tik 42% tikslumu. Modelio tikslumo gerinimui pasitelkus sukurtą krepšinio rungtynių duomenų rinkinį, rezultatai pagerėjo iki 82%. Pastebėta, kad didžiausią įtaką tikslumui daro prekių ženklų reklamų persidengimas bei kameros judesio sukeltas išblukimas.
5. Atlikus aptiktų objektų rezultatų apdorojimą pastebėta, kad vaizdinių savybių gavimo modelio pasirinkimas daro didelę įtaką prekių ženklų grupavimo rezultatų tikslumui. Nustatyta, kad naudojant VGG gaunami 3 % tikslesni grupavimo rezultatai nei *MobileNet*, bei 10 % tikslesni nei naudojant *DenseNet19*. Visgi, šiame uždavinyje dideliu iššūkiu išlieka neteisingai aptikti prekių ženklai.

## Literatūros sąrašas

- [1] Event Marketing Institute, „Experiential Marketing Content Benchmarking Report,“ Event Marketing Institute, 2016.
- [2] R. Gupta ir A. Gandhi, „Impact of Elements of Ad's on Sports Fan Attitude during a Live Sporting Event,“ *Cardiometry*, t. 24, pp. 867-876, 2022.
- [3] S. Ishak ir N. F. M. Zabil, „Impact of consumer awareness and knowledge to consumer effective behavior,“ *Asian Social Science*, t. 8, nr. 13, p. 108.
- [4] S. Hou, J. Li, W. Min, Q. Hou, Y. Zhao, Y. Zheng ir S. Jiang, „Deep learning for logo detection: A survey,“ *ACM Transactions on Multimedia Computing, Communications and Applications*, t. 20, nr. 3, pp. 1-23, 2023.
- [5] J. Jacobs, P. Jain ir K. Surana, „Is sports sponsorship worth it?,“ McKinsey & Company, 2014.
- [6] B2B International, „How to Measure Sponsorship Effectiveness and ROI“.
- [7] A. Mahmoudian, S. S. Boroujerdi, S. Mohammadi ir V. Delshab, „Testing the impact of athlete brand image attributes on fan loyalty,“ 2020.
- [8] J. A. Fortunato, *Sports Sponsorship: Principles and Practices*, McFarland & Company, 2013.
- [9] J. Plush, What is Sponsor Media Value (SMV)?, Sponsorship Innovation, relo metrics, 2023.
- [10] Y. Kalantidis, L. G. Pueyo, M. Trevisiol, Z. Roelof van ir Y. Avrithis, „Scalable triangulation-based logo recognition,“ *ACM International Conference on Multimedia Retrieval*, 2011.
- [11] S. Romberg, L. G. Pueyo, R. Lienhart ir R. Van Zwol.
- [12] A. Joly ir O. Buisson, „Logo retrieval with a contrario visual query expansion,“ *Proceedings of the 17th ACM international conference on Multimedia*, pp. 581-584.
- [13] A. Kuznetsov ir A. V. & Savchenko, „A new sport teams logo dataset for detection tasks,“ *Computer Vision and Graphics: International Conference*, pp. 87-97, 2020.
- [14] H. Su, X. Zhu ir S. Gong, „Open logo detection challenge,“ *arXiv preprint arXiv*, 2018.
- [15] Q. Hou, W. Min, J. Wang, S. Hou, Y. Zheng ir S. Jiang, „Foodlogodet-1500: A dataset for large-scale food logo detection via multi-scale feature decoupling network,“ *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 4670-4679, 2021.
- [16] J. Wang, W. Min, S. Hou, S. Ma, Y. Zheng ir S. Jiang, „Logodet-3k: A large-scale image dataset for logo detection,“ *ACM Transactions on Multimedia Computing, Communications, and Applications*, t. 18, nr. 1, pp. 1-19, 2022.
- [17] K. O'Shea ir R. Nash, „An Introduction to Convolutional Neural Networks,“ *arXiv*, 2015.
- [18] S. Targ, D. Almeida ir K. Lyman, „RESNET IN RESNET: GENERALIZING RESIDUAL ARCHITECTURES,“ t. 1603.08029v1, 2016.
- [19] W. Wong, „What is Residual Connection? A technique for training very deep neural networks,“ *Towards Data Science*, 2021.
- [20] Y. Zhu ir S. Newsam, „Densenet for dense flow,“ *IEEE international conference on image processing*, pp. 790-794, 2017.
- [21] G. Huang, Z. Liu ir M. Laurens van der, „Densely Connected Convolutional Networks,“ *Computer Vision Foundation, CVPR*, 2017.



- [22] S. Feng, L. Zhao, H. Shi, M. Wang, S. Shen ir W. Wang, „One-dimensional VGGNet for high-dimensional data.,“ *Applied Soft Computing*, t. 135, 2023.
- [23] S. Tammina, „Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images,“ *International Journal of Scientific and Research Publications*, t. 9, nr. 10, pp. 143 - 149, 2019, Spalis.
- [24] R. Girshick, J. Donahue, T. Darrell ir J. Malik, „Region-based convolutional networks for accurate object detection and segmentation,“ *IEEE transactions on pattern analysis and machine intelligence*, t. 38, nr. 1, pp. 142-158, 2015.
- [25] S. M. Abbas ir N. ., S. Shailendra, „Region-based Object Detection and Classification,“ ĩtraukta *Computational Intelligence and Communication Technology*, 2018.
- [26] R. Girshick, „Fast R-CNN,“ *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440 - 1448, 2015.
- [27] R. Shaoqing, H. Kaiming, G. Ross ir S. Jian, „Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,“ *NeuralPS Proceedings*, 2015.
- [28] J. Redmon, S. Divvala, R. Girshick ir A. Farhadi, „You Only Look Once: Unified, Real-Time Object Detection,“ arXiv, 2016.
- [29] D. Juan, „Understanding of Object Detection Based on CNN Family and YOLO,“ *Journal of Physics: Conference Series*, t. 1004, 2018.
- [30] J. Redmon ir A. Farhadi, „YOLOv3: An Incremental Improvement,“ arXiv, Cornell University, 2018.
- [31] W. Budiharto, E. Abdulrachman, Y. Heryadi ir Sarwo, „Logo detection and brand recognition with one-stage logo detection framework and simplified resnet50 backbone,“ IEEE, Yogyakarta, Indonesia, 2019.
- [32] A. Alsheikhy, S. Yahia ir M. Barr, „Logo Recognition with the Use of Deep Convolutional Neural Networks,“ *Engineering, Technology & Applied Science Research*, %1 t. iř %210, No. 5, pp. 6191-6194, 2020.
- [33] Y. Bao, H. Li, X. Fan, R. Liu ir Q. Jia, „Region-based CNN for Logo Detection,“ *ICIMCS*, 2016.
- [34] E. Ashna, F. Tony ir C. V. Hari, „International Conference on Advanced Computer Control, ICACC,“ ĩtraukta *YOLO based Logo detection*, 2019.
- [35] Z. Lu, J. Lu, Q. Ge ir T. Zhan, „Multi-object Detection Method based on YOLO and ResNet Hybrid Networks,“ ĩtraukta *4th International Conference on Advanced Robotics and Mechatronics*, 2019.
- [36] M. Ahmed, R. Seraj ir S. M. S. Islam, „The k-means algorithm: A comprehensive survey and performance evaluation.,“ *Electronics*, t. 9, nr. 8, 2020.
- [37] D. Deng, „DBSCAN clustering algorithm based on density,“ *7th international forum on electrical engineering and automation*, pp. 949-953, 2020.
- [38] K. P. Sinaga ir Y. Miin-Shen, „Unsupervised K-means clustering algorithm.,“ *EEE access* 8, pp. 80716-80727, 2020.
- [39] M. Caron, P. Bojanowski, A. Joulin ir M. Douze, „Deep clustering for unsupervised learning of visual features,“ *Proceedings of the European conference on computer vision* , pp. 132-149, 2018.

- [40] X. Jiang, K. Sun, L. Ma, Z. Qu ir C. Ren, „Vehicle logo detection method based on improved YOLOv4,“ *Electronics*, t. 11, nr. 20, 2022.
- [41] S. Yang, J. Zhang, C. Bo, M. Wang ir L. Chen, „Fast vehicle logo detection in complex scenes,“ *Optics & Laser Technology*, t. 110, pp. 196-201.
- [42] J. Zhang, S. Yang, C. Bo ir Z. Zhang, „ehicle logo detection based on deep convolutional networks,“ *Computers & Electrical Engineering*, t. 90.
- [43] C. Wilms, R. Heid, M. A. Sadeghi, A. Ribbrock ir S. Frintrop, „Which airline is this? Airline logo detection in real-world weather conditions,“ *25th International Conference on Pattern Recognition (ICPR)*, pp. 4996-5003, 2020.
- [44] Jain, R. Kumar ir e. al, „LogoNet: Layer-aggregated attention centernet for logo detection,“ *IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1-6, 2021.
- [45] C. Wilms ir S. & Frintrop, „AttentionMask: Attentive, efficient object proposal generation focusing on small objects,“ *4th Asian Conference on Computer Vision. Computer Vision–ACCV*, pp. 678-694, 2018.
- [46] G. Zhu ir D. & Doermann, „Automatic document logo detection,“ *Ninth International Conference on Document Analysis and Recognition*, t. 2, pp. 864-868.
- [47] F. Leng, „A gradient balancing approach for robust logo detection,“ *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 4765-4769, 2021.
- [48] Y. Luo, J. Zhu ir T. Pfister, „A A simple yet effective baseline for robust deep learning with noisy labels,“ *arXiv preprint arXiv*, 2019.
- [49] H. Su, S. Gong ir X. Zhu, „Multi-perspective cross-class domain adaptation for open logo detection,“ *Computer vision and image understanding*, t. 204, 2021.
- [50] A. Tüzkö, C. Herrmann, D. Manger ir J. Beyerer, „Open set logo detection and retrieval,“ *arXiv preprint arXiv*, 2017.
- [51] J. Solawetz, „What is YOLOv7? A Complete Guide,“ Roboflow, 17 7 2022. [Tinkle]. Available: <https://blog.roboflow.com/yolov7-breakdown/>.
- [52] A. KHAZRI, „Faster RCNN Object detection,“ Towards Data Science, 2019.
- [53] A. Joly ir O. Buisson, „Logo retrieval with a contrario visual query expansion,“ *ACM international conference of Multimedia*, 2009.
- [54] A. Joly ir O. Buisson, „Logo retrieval with a contrario visual query,“ *ACM international conference on Multimedia*, 2009.
- [55] S. C. Hoi, X. Wu, H. Liu, Y. Wu, H. Wang, H. Xue ir Q. Wu, „LOGO-Net: Large-scale Deep Logo Detection and Brand Recognition,“ School of Information Systems, Singapore Management University, Singapore, Hangzhou, China.
- [56] Y. Kalantidis, L. G. Pueyo, M. Trevisiol, R. Van Zwol ir Y. Avrithis, „Scalable triangulation-based logo recognition,“ *Proceedings of the 1st ACM international conference on multimedia retrieval*, pp. 1-7.

## Providing brands visibility data in live sports videos using deep learning algorithms

Julius Gudauskas<sup>a</sup>

<sup>a</sup> Department of Applied Informatics, Kaunas University of Technology, Studentų 50, Kaunas, Lithuania  
[julius.gudauskas@ktu.edu](mailto:julius.gudauskas@ktu.edu)

### Abstract

In the dynamic landscape of marketing and advertising, assessing brand visibility in live sports events plays a pivotal role in understanding brand exposure and impact. Traditional methods of manual annotation and analysis are time-consuming and subjective, necessitating automated solutions for efficient and objective evaluation. In this study proposed a novel approach leveraging deep learning algorithms to evaluate brand visibility in live sports videos. This research employs state-of-the-art object detection models, such as YOLO (You Only Look Once) and Faster R-CNN, to detect and localize brand logos within video frames. By training these models on annotated open-source logo datasets, we can extract valuable insights about the brands. The experimental results demonstrate the effectiveness of the proposed methodology in detecting logos and providing a valuable data about the positions for brand owners.

### Keywords 1

Brands visibility, logo detection, YOLO, Faster R-CNN

## 1. Introduction

In the world of sports live video streaming, a considerable number of brands are trying to get noticed by the audience using various popular visibility materials: posters, stickers, billboards, etc. The research evaluating the impact and effectiveness of advertisements in sports arenas emphasizes that people notice at least some of the advertisements they are exposed to and usually remembers a part of them that were the most noticeable [1]. Typically, clients engage in negotiations with advertising executives to determine the conditions that will govern brand placement in the arena – specifying factors like coverage, frequency of display on advertisement billboards, and overall visibility strategies. A study conducted by Eventmarketer [2] found that 72% of the audience are captivated by the brand when they see it during the events like music festivals or sports competitions where they are provided with good emotions and excitement. These occasions, characterized by heightened emotions and excitement, offer a unique opportunity for brands to establish a connection with a vast and diverse audience, potentially converting them into new users. However, to reach a wider audience, the brand must be placed in a visible location. Studies has shown that locations, such as boundary line hoarding are considered as the perfect place to display brand logos without irritating the viewers and getting maximum visibility [3] [4]. As sponsorship agreements comes with a

---

Providing brands visibility data in live sports videos using deep learning algorithms, March 25<sup>th</sup>, 2024, Kaunas, Lithuania

EMAIL: [julius.gudauskas@ktu.edu](mailto:julius.gudauskas@ktu.edu)

ORCID: 0009-0009-6351-5080



© 2024 Information Society and University Studies  
2024, May 17, Kaunas, Lithuania. Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)