

VILNIAUS UNIVERSITETAS
INFORMATIKOS INSTITUTAS
PROGRAMŲ SISTEMŲ KATEDRA

**Automobilių numerių atpažinimas naudojant Tesseract
LSTM rekurentinį neuroninį tinklą**

Car Number Plate Recognition Using Tesseract LSTM Recurrent Neural
Network

Bakalauro baigiamasis darbas

Atliko: Emilis Ruzveltas (parašas)

Darbo vadovas: dr. Vytautas Valaitis (parašas)

Recenzentas: lekt. Tomas Smagurauskas (parašas)

Vilnius
2019

Santrauka

Summary

TURINYS

ĮVADAS	4
1. Duomenų generavimas	7
1.1. Rėmelio atpažinimui skirtų duomenų generavimas	7
1.1.1. Pirminis duomenų generavimo variantas	7
1.1.1.1. Generavimo parametrai	7
1.1.1.2. Numerio generavimas	9
1.1.1.3. Transformacija	9
1.1.1.4. Kompozicija	11
1.1.1.5. Triukšmo pridėjimas	12
1.1.2. Duomenų generavimo patobulinimai	12
1.2. Numerio atpažinimui skirtų duomenų generavimas	12
1.2.1. Idėja	13
1.2.2. Simbolių rinkimas	13
1.2.3. Generavimo algoritmas	13
2. Neuroninių tinklų architektūra	14
2.1. Numerio rėmelio atpažinimui skirtas neuroninis tinklas	14
2.1.1. Modelis	14
2.2. Numerio simbolių atpažinimui skirtas neuroninis tinklas	17
2.2.1. Bendrai apie LSTM	17
2.2.1.1. Rekurentinis neuroninis tinklas	18
2.2.2. Integracija su Tesseract	19
2.2.3. Sisteminiai reikalavimai	19
2.2.4. Įgyvendinimo pagrindai	19
2.2.5. Naujo tinklo lygio pridėjimas	19
2.2.6. VGSL specifikacijos	20
2.2.7. Vidinių tinklo lygių sintaksė	21
2.2.8. Kintamo dydžio įvestis ir apibendrinantis LSTM lygis	24
2.2.9. Modelis	24
3. Neuroninių tinklų apmokymas	26
3.1. Konvoliucinio neuroninio tinklo mokymas	26
3.2. LSTM rekurentinio neuroninio tinklo mokymas	27
3.2.1. Atpažinimo kokybės gerinimas	28
3.2.1.1. Paveikslėlio apdorojimas	28
3.2.1.2. Puslapių skirstymo metodas	33
3.2.1.3. Žodynai, žodžių sąrašai, šablonai	34
3.2.2. Rinkmenų pasiruošimas	34
4. Vaizdo atpažinimas	36
4.1. Numerio rėmelio atpažinimas	36
4.2. Numerio simbolių atpažinimas	37
REZULTATAI	38
IŠVADOS	38
LITERATŪROS SĄRAŠAS	39
SĄVOKŲ APIBRĖŽIMAI	40
SANTRUMPOS	40

Įvadas

Pagrindinis automobilių atpažinimo sistemų tikslas yra automatizuoti vaizdo stebėjimą ir apdorojimą bei automatiškai surinkti įvairią informaciją apie transporto priemonę. Automobilių numerių atpažinimo sistemos remiasi tuo, kad kiekviena transporto priemonė turi unikalų identifikacinį kodą, kuris leidžia vienareikšmiškai nustatyti transporto priemonės savininką. Techniškai automobilių numerių atpažinimas yra paveikslėlių apdorojimo programa, naudojantis specialiu algoritmu išgauti rezultatus iš paveikslėlio. Automatinis paveikslėlių atpažinimas turi platų spektrą pritaikymo sričių, tokių kaip automobilių patikra, automatinis kelių mokesčių surinkimas, išmanus eismo reguliavimas [BSS13]. Didžioji dauguma automobilio numerių atpažinimo sistemų remiasi optine ženklų atpažinimo sistema. Jų apdorojimo greitis yra pakankamai greitas, kad būtų efektyviai išnaudojama įvairiose srityse. Tačiau dažniausiai yra kuriamos specializuotos atpažinimo programos skirtingiems regionams. Panaudojus neuroninius tinklus galima būtų apmokyti atpažinti numerius, kurių formatai yra skirtingi. Taip pat galima būtų pagreitinti procesą iškerpant numerio rėmelį iš paveikslėlio pasinaudojus neuroniniais tinklais. Norint pagreitinti patį teksto atpažinimą galima naudoti rekurentinį neuroninį tinklą su LSTM savybėmis [LS16]. Šiame darbe naudosime kursinio darbo metu sukurtą konvoliucinį neuroninį tinklą, kuris yra skirtas atpažinti numerio rėmelio koordinatas. Taip pat pritaikysime bei modifikuosime Tesseract LSTM rekurentinį neuroninį tinklą, kuris sugebės atpažinti automobilio numerio simbolius [Smi07].

Darbo tikslas

Sukurti programą, kuri gebėtų atpažinti lietuviškus automobilio numerius paveikslėlyje panaudojant Tesseract LSTM rekurentinį neuroninį tinklą.

Uždaviniai

--TODO

1. Pasinaudojus paveikslėlių duomenų rinkiniu susigeneruoti 1.000.000 atsitiktinių paveikslėlių su automobilio numeriais.
2. Apmokyti kursinio darbo metu sukurtą konvoliucinį neuroninį tinklą (rėmelio atpažinimui) pateikiant sugeneruotus paveikslėlius.

3. Apmokyti Tesseract LSTM rekurentinį neuroninį tinklą (teksto atpažinimui) pateikiant sugeneruotus paveikslėlius.
4. Pasinaudojus kursinio darbo metu sukurtu ir apmokytu konvoliuciniu neuroniniu tinklu atpažinti numerio rėmelį paveikslėlyje ir gauti jo koordinates.
5. Pagal gautas koordinates, iškirpti rėmelį ir pasinaudojus Tesseract LSTM neuroniniu tinklu atpažinti numerį bei atvaizduoti gautus rezultatus pradiniam paveikslėlyje.
6. Ištestuoti tinklą su tikrais paveikslėliais, kuriuose yra lietuviški automobilių numeriai.

Darbo prielaidos ir metodika

Šiais laikais, kai dominuoja naujosios technologijos, paremtos dirbtiniu intelektu, svarbu analizuoti ir gilintis į procesus, kurie nusako kaip veikia neuroniniai tinklai. Analizuojant bei tobulinant dirbtinio intelekto sistemas, galima pasiekti greitesnių bei efektyvesnių rezultatų nei naudojant tradicinius atpažinimo metodus.

Tyrimo objektas yra automobilių numerių atpažinimas. Bus tiriama kaip vyksta teksto atpažinimas pasitelkiant dirbtinius neuroninius tinklus.

Pagrindinis šio tyrimo metodas – rekurentinio LSTM dirbtinio neuroninio tinklo veikimo analizė. Analizuojama pasitelkiant įvairius mokslinius šaltinius, straipsnius, publikacijas, knygas. Kitoje darbo dalyje bus atliekamas eksperimentas pritaikant teoriją.

Darbo atlikimo procesas

Pirmiausia bus gilinamasi į Tesseract LSTM neuroninio tinklo veikimo principus [BSS13]. Išanalizavus, bus bandoma apmokyti neuroninį tinklą su kursinio darbo metu sugeneruotais paveikslėliais. Atlikus apmokymą, reikės analizuoti ir gerinti tikslumą keičiant neuroninio tinklo specifikaciją. Galiausiai norint pasiekti dar didesnę spartą ir tikslumą, tinklas bus pritaikytas atpažinti lietuviškus automobilio numerius. Atlikus šį eksperimentą bus sukurta programa, kuri naudos kursinio darbo metu sukurtą konvoliucinį neuroninį tinklą skirtą atpažinti rėmelį bei šiame darbe sukurtą bei modifikuotą Tesseract LSTM neuroninio tinklo konfigūraciją.

Eksperimente naudojami instrumentai

Atlikti šiam eksperimentui buvo naudojami šie pagrindiniai įrankiai:

- Tesseract – skirta atpažinti numeryje esančius simbolius.
- Python – programavimo kalba naudota kurti programoms.
- TensorFlow – skirta neuroninio tinklo pagalba atpažinti numerio rėmelį nuotraukoje.
- OpenCV – skirta apdoroti paveikslėlius.

1. Duomenų generavimas

1.1. Rėmelio atpažinimui skirtų duomenų generavimas

Norint sukurti realiai veikiančią programą, kuri naudotų neuroninį tinklą išgauti tikėtinam rezultatui, tinklą reikia apmokyti su dideliu kiekiu duomenų. Apmokant bet kokį neuroninį tinklą turi būti pateiktas duomenų rinkinys su norimu gauti rezultatu.

1.1.1. Pirminis duomenų generavimo variantas

Pirminis duomenų generavimo variantas, kuris buvo įgyvendintas bei sėkmingai sugeneruoti 100.000 paveikslėlių skirtų neuroninio tinklo apmokymui.

1.1.1.1. Generavimo parametrai

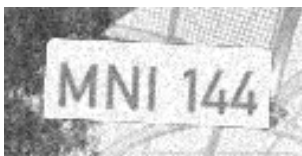
Šiam tyrimui buvo pasirinkta generuoti duomenų rinkinį, kurių kiekvienas paveikslėlis būtų 128 pikselių pločio ir 64 pikselių ilgio. Toks pasirinktas būdas užtikrina, kad neuroninis tinklas bus pajėgus suprasti paveikslėlio turinį, o dydis pakankamai mažas, kad būtų galima turėti efektyviai veikiančią neuroninį tinklą.

Pirmoji paveikslėlio rezultato dalis nurodo, koks yra teisingas numeris. Antroji – numerio rėmelio egzistavimas paveikslėlyje. Jei reikšmė 1 – numeris yra tinkamas nuskaitymui, 0 – neatitinka kriterijų (2 pav.).

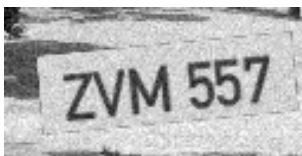
Kriterijai, kurie nusako ar numerio rėmelis yra tinkamas apdorojimui:

- Visas rėmelio plotas yra paveikslėlyje.
- Rėmelio plotis yra mažesnis nei 80% paveikslėlio pločio.
- Rėmelio aukštis yra mažesnis nei 87,5% paveikslėlio aukščio.
- Rėmelio plotis yra didesnis nei 60% paveikslėlio pločio.
- Rėmelio aukštis yra didesnis nei 60% paveikslėlio aukščio.

Su tokiais parametrais galima naudoti judantį 128x64 pikselių langelį, kuris judėtų po 8 pikselius ir kas kartą padidintų rėmelį $\sqrt{2}$ kartų. Tokiu būdu užtikrinama, kad nebus praleista nei viena paveikslėlio vieta, o taip pat pakankamai efektyviai ir greitai pereinamas visas paveikslėlis.



(a) Tikimasis rezultatas **MNI144 1**.



(b) Tikimasis rezultatas **ZVM557 1**.



(c) Tikimasis rezultatas **COH150 0** (per mažas numeris).



(d) Tikimasis rezultatas **GTK311 0** (ne pilnas numeris).

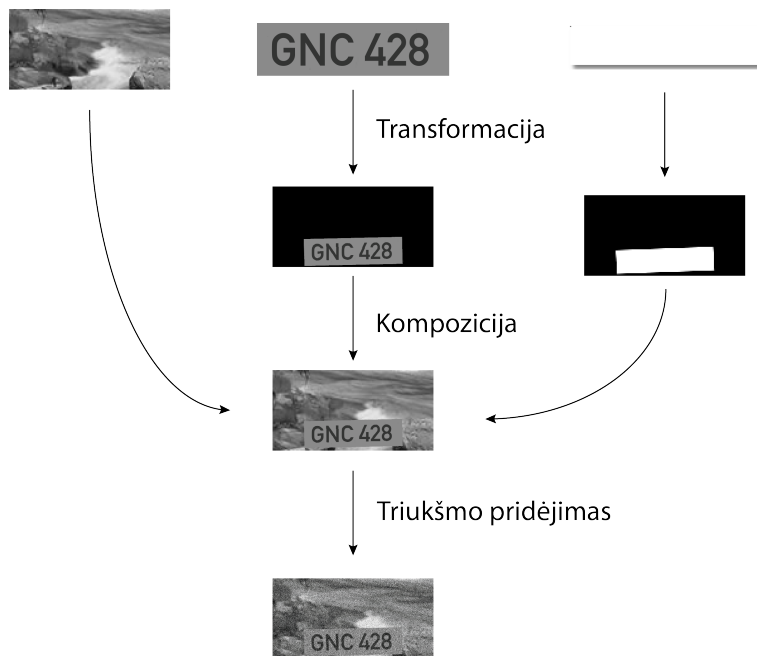


(e) Tikimasis rezultatas **ABN046 0** (ne pilnas numeris).



(f) Tikimasis rezultatas **KLF155 0** (nėra numerio).

2 pav. Sugeneruotų paveikslėlių pavyzdžiai



3 pav. Paveikslėlių generavimo schema

1.1.1.2. Numerio generavimas

Numeris ir rėmelio spalva generuojama atsitiktinai, tačiau tekstas turi būti tamsesnis negu rėmelis. Tokiu būdu bandoma atkurti realaus pasaulio apšvietimo variacijas. Numeris generuojamas pagal Lietuvos Respublikos Valstybinių numerių formatą, kuris yra - 3 lotyniško alfabeto raidės (išskyrus lietuvių kalboje nenaudojamas raides) ir 3 arabiški skaičiai. Generuojant numerį, atsitiktine tvarka parenkamos trys raidės iš 23 raidžių žodyno *ABCDEFGHIJKLMNPRSTUVYZ* bei 3 skaičiai iš skaičių žodyno *0123456789*. Maksimalus galimas unikalių numerių skaičius siekia:

$$23^3 * 10^3 = 12.167.000.$$

1.1.1.3. Transformacija

Norint, kad tinklas efektyviai mokytųsi ir atpažintų paveikslėlius realaus pasaulio sąlygomis, generuojant duomenų rinkinį buvo pritaikyta rėmelio transformacija. Tai atlikti buvo pasitelktas metodas generuoti atsitiktines reikšmes X , Y , Z ašims ir pritaikyti Oilerio kampų metodą[Sla99]. Reikšmių režiai pasirinkti tokie, kuriuos labiausiai tikėtina sutikti realiame pasaulyje. Kaip atrodo transformacija galima matyti 3 paveikslėlyje. Ašių atsitiktinių reikšmių režiai:

$$-0.3 \leq X \leq 0.3,$$

$$-0.2 \leq Y \leq 0.2,$$

$$-1.2 \leq Z \leq 1.2.$$

Transformacijos vykdomos 3 etapais (programinis kodas matomas 4 pav.):

1. Sukama aplink Y ašį:

- Apskaičiuojamos $\cos(Y)$ ir $\sin(y)$ reikšmės,
- Sudaroma 3×3 matrica su reikšmėmis,

$$\begin{bmatrix} c & 0 & s \\ 0 & 1 & 0 \\ -s & 0 & c \end{bmatrix}.$$

2. Sukama aplink X ašį

- Apskaičiuojamos $\cos(X)$ ir $\sin(X)$ reikšmės,
- Sudaroma 3x3 matrica su reikšmėmis,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & c & -s \\ 0 & s & c \end{bmatrix}.$$

- Sudauginama su praeitame žingsnyje gauta matrica

3. Sukama aplink Z ašį

- Apskaičiuojamos $\cos(Z)$ ir $\sin(Z)$ reikšmės,
- Sudaroma 3x3 matrica su reikšmėmis,

$$\begin{bmatrix} c & 0 & s \\ 0 & 1 & 0 \\ -s & 0 & c \end{bmatrix},$$

- Sudauginama su praeitame žingsnyje gauta matrica.

```
def euler_to_mat(yaw, pitch, roll):
    # Rotate clockwise about the Y-axis
    c, s = math.cos(yaw), math.sin(yaw)
    M = numpy.matrix([[ c, 0., s],
                      [ 0., 1., 0.],
                      [-s, 0., c]])

    # Rotate clockwise about the X-axis
    c, s = math.cos(pitch), math.sin(pitch)
    M = numpy.matrix([[ 1., 0., 0.],
                      [ 0., c, -s],
                      [ 0., s, c]]) * M

    # Rotate clockwise about the Z-axis
    c, s = math.cos(roll), math.sin(roll)
    M = numpy.matrix([[ c, -s, 0.],
                      [ s, c, 0.],
                      [ 0., 0., 1.]]) * M

    return M
```

4 pav. Oilerio kampų metodo kodas

1.1.1.4. Kompozicija

Turėti realų foną svarbu, kadangi tinklas turi išmokti surasti rėmelio kampus „nesukčiaudamas“. Naudojant juodą foną, tinklas gali daryti prielaidą, kad rėmelis yra ten, kur nėra juodos spalvos, o tai būtų netikslu realiame pasaulyje. Transformuotas automobilio numerio rėmelis sukomponuojamas su atsitiktiniu paveikslėliu atsitiktinėje vietoje. Atsitiktinių paveikslėlių šaltiniu naudojamas daugiau nei 100.000 paveikslėlių duomenų rinkinys [XHE⁺10]. Labai svarbu didelis kiekis paveikslėlių, taip sumažinant riziką, kad neuroninis tinklas atsimins kiekvieną paveikslėlį. Kaip atrodo kompozicija galima matyti 3 paveikslėlyje.

$$A = 0.6,$$

$$B = 0.875,$$

$$C = 1.5.$$

kur:

- A - minimalus numerio rėmelio plotis,
- B - maksimalus numerio rėmelio aukštis,
- C - dydžio variacijos koeficientas.

$$\min = (A + B) * 0.5 - (B - A) * 0.5 * C,$$

$$\min = ((0.6 + 0.875) * 0.5) - ((0.875 - 0.6) * 0.5 * 1.5),$$

$$\min = (1.475 * 0.5) - (0.275 * 0.5 * 1.5),$$

$$\min = 0.7375 - 0.20625,$$

$$\mathbf{\min = 0.53125},$$

$$\max = (A + B) * 0.5 + (B - A) * 0.5 * C,$$

$$\max = ((0.6 + 0.875) * 0.5) + ((0.875 - 0.6) * 0.5 * 1.5),$$

$$\max = (1.475 * 0.5) + (0.275 * 0.5 * 1.5),$$

$$\max = 0.7375 + 0.20625,$$

$$\mathbf{\max = 0.94375},$$

$$x = R[\min, \max],$$

$$p = 1, \text{ kai } x \in [A, B],$$

$$p = 0, \text{ kai } x \in [A, B].$$

kur:

- min - minimalus generuojamas dydžio koeficientas,
- max - maksimalus generuojamas dydžio koeficientas,
- R - atsitiktinio skaičiaus generavimo funkcija tarp dviejų reikšmių,
- x - numerio rėmelio dydžio koeficientas lyginant su pradiniu dydžiu,
- p - jei 1 - rėmelis tinkamai egzistuoja paveikslėlyje, 0 - rėmelis neegzistuoja arba yra netinkamas.

Atsitiktinių reikšmių rėžis, kuris nusako kurioje vietoje turėtų atsidurti rėmelis.

1.1.1.5. Triukšmo pridėjimas

Triukšmas paveikslėlyje reikalingas, kadangi realiame pasaulyje pasitaiko, kad kameros sensorius generuoja triukšmus, o taip pat, kad neuroninis tinklas nepersimokytų ir neskirstytų paveikslėlių pagal vieną konkrečią spalvą ar būtų priklausomas nuo „aštrių“ kampų. Triukšmas paveikslėliui pridedamas pritaikant Gauso normalųjį skirstinį su reikšme 0.05. Kaip atrodo triukšmo pridėjimas galima matyti 3 paveikslėlyje.

1.1.2. Duomenų generavimo patobulinimai

--TODO 1h kas patobulinta remeliams kaip pasikeitė statistika

1.2. Numerio atpažinimui skirtų duomenų generavimas

Norint sėkmingai apmokyti neuroninį tinklą, reikia daug pradinių mokymo duomenų. Šiam tikslui pasiekti nuspręsta duomenis susigeneruoti, kadangi tiek daug lietuviškų numerių tikrų nuotraukų nėra įmanoma gauti.

1.2.1. Idėja

Norint kuo efektyviau apmokyti LSTM rekurentinį neuroninį tinklą, reikia sugeneruoti tokio pačio šrifto atsitiktinius numerius, kurie kuo panašiau atkurtų realią situaciją. Idėja buvo susirinkti visas galimas raides ir skaičius iš realių automobilių nuotraukų. Atrinktus simbolius išsikirti bei sugeneruoti atsitiktinius raidžių ir skaičių kratinius.

1.2.2. Simbolių rinkimas

Realijų automobilių numerių nuotraukų paieška buvo vykdoma <http://autoplius.lt> puslapyje. Norint iškirpti kokybiškas raides bei skaičius reikia aukštos kokybės nuotraukų. Atrinkus tinkamas nuotraukas, buvo iškirptos visos galimos raidės ir skaičiai (5 pav.). Kai kuriems simboliams reikėjo pritaikyti transformacijas, kad jų orientacija būtų horizontaliai tiesi.

ABCDEFGHIJKLMN~~OP~~RSTUVZ0123456789

5 pav. Atrinktos raidės ir skaičiai

1.2.3. Generavimo algoritmas

Generuoti atsitiktiniams automobilio numeriams parašyta Python programėlė. Veikimo eiga:

- Masyve *letters* saugomi paveikslėliai atitinkantys raides.
- Masyve *numbers* saugomi paveikslėliai atitinkantys skaičius.
- Sukamas ciklas N kartų.
- Kiekvieno iteracijos metu atsitiktiniu būdu atrenkamos trys raidės ir trys skaičiai iš atitinkamo masyvo.
- Sukuriamas naujas masyvas, kuriame iš eilės sudedami atrinkti paveikslėliai.
- Surandamas mažiausias paveikslėlis iš atrinktų, ir pagal jo dydį sumažinamas likusių paveikslėlių aukštis proporcingai.
- Sujungiamas vienas paveikslėlis iš atrinktų simbolių.
- Paveikslėlis išsaugomas *xxxxyy.tif* formatu, kur *x* – raidė, *y* – skaičius.
- Šalia išsaugomas tekstinė rinkmena *xxxxyy.gt.txt*, kur *x* – raidė, *y* – skaičius, kurio viduje yra XXXYYY formatu išsaugotas sugeneruotas numeris.

2. Neuroninių tinklų architektūra

Šiame skyriuje aprašyti dviejų skirtingų neuroninių tinklų architektūriniai sprendimai. Numerio rėmelio atpažinimui naudojamas konvoliucinis neuroninis tinklas bei numeryje esančių simbolių atpažinimui naudojamas rekurentinis neuroninis tinklas.

2.1. Numerio rėmelio atpažinimui skirtas neuroninis tinklas

Numerio rėmelio atpažinimui panaudotas konvoliucinis neuroninis tinklas.

2.1.1. Modelis

Neuroninis tinklas kurtas su Tensorflow bibliotekomis. Kuriant dirbtinį konvoliucinį neuroninį tinklą buvo pasirinkta architektūra pavaizuota 6 pav. Iš viso yra 3 konvoliuciniai lygiai, kurių dydžiai yra 48, 64 ir 128[GBI⁺13]. Visų jų langelio dydis yra vienodas – 5x5. Taip pat yra 3 max pool'ingo lygmenys, kurių pirmo ir trečio langelio dydis yra 2x2, o antro – 1x2. Tada neuroninis tinklas turi du pilnai sujungtus lygius, kurių pirmojo dydis – 2048, o antrojo (klasifikatoriaus) – 1. Po pirmojo konvoliucinio lygio pritaikyta neuronų atmetimo operacija, norint nepermokyti tinklo pirminėje stadijoje. Po trečiojo konvoliucinio lygio taip pat pritaikyta neuronų atmetimo operacija, norint padidinti neuroninio tinklo tikslumą, kadangi pastebėta, kad ignoruojant 50% neuronų, tinklas turi didesnę atpažinimo tikslumą[SHT⁺15]. Kiekvieno mokymo ciklo metu imties dydis yra 50. Galutinis tinklo išvedamas rezultatas yra:

$$0 \leq x \leq 1, x \in N.$$

Neuroninį tinklą sudaro:

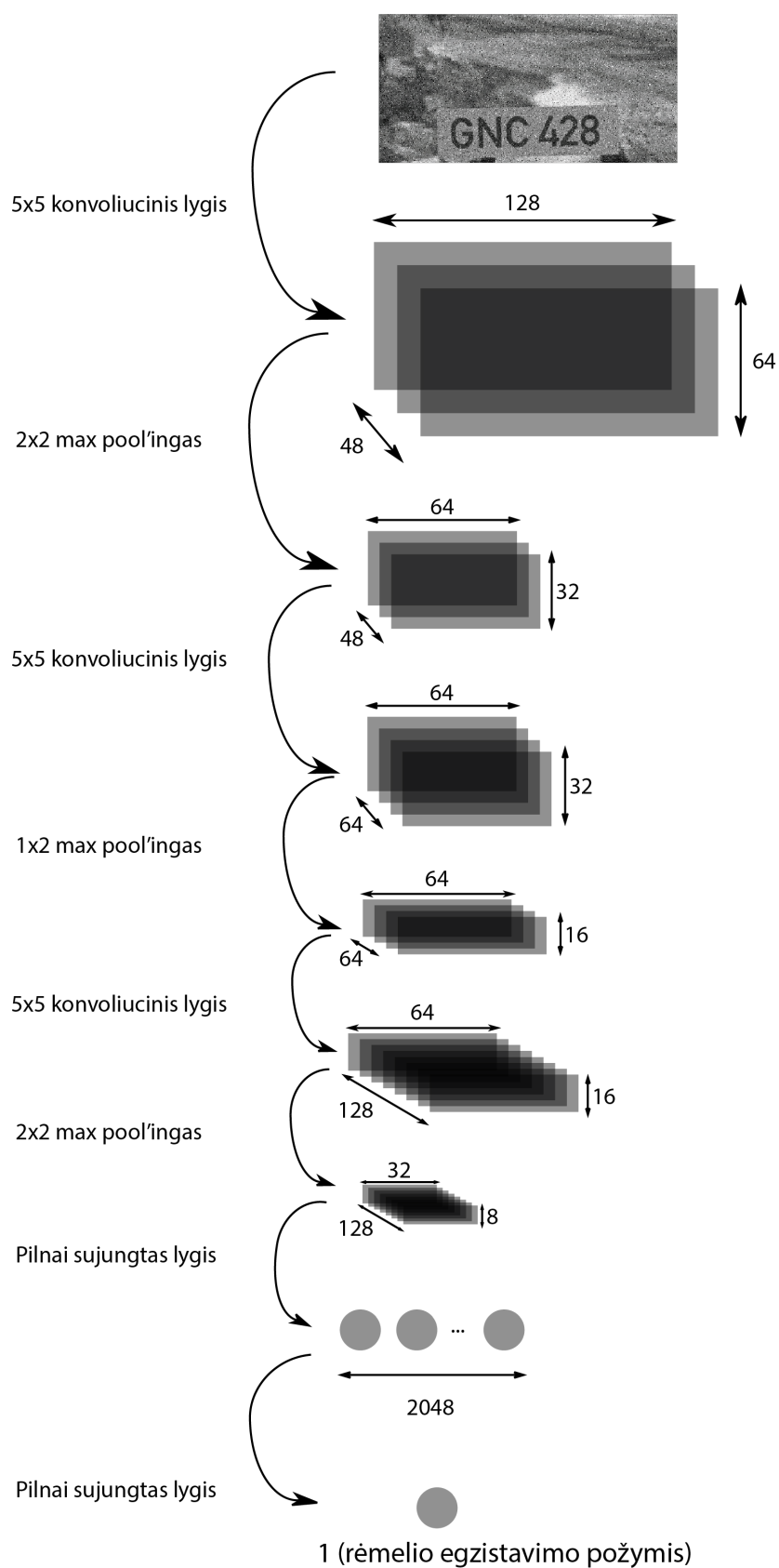
- 3 konvoliuciniai lygiai:
 1. Konvoliucinis lygis – 48 filtrų, langelio dydis 5x5, įeinančio paveikslėlio dimensijos 128x64x3, išeinančio paveikslėlio dimensijos 128x64x48.
 2. Konvoliucinis lygis – 64 filtrų, langelio dydis 5x5, įeinančio paveikslėlio dimensijos 64x32x48, išeinančio paveikslėlio dimensijos 64x32x64.
 3. Konvoliucinis lygis – 128 filtrų, langelio dydis 5x5, įeinančio paveikslėlio dimensijos 64x16x64, išeinančio paveikslėlio dimensijos 64x16x128.

- 3 max pool'ingo lygiai:

1. Max pool'ingo lygis – langelio dydis 2×2 , įeinančio paveikslėlio dimensijos $128 \times 64 \times 48$, išeinančio paveikslėlio dimensijos $64 \times 32 \times 48$.
2. Max pool'ingo lygis – langelio dydis 1×2 , įeinančio paveikslėlio dimensijos $64 \times 32 \times 64$, išeinančio paveikslėlio dimensijos $64 \times 16 \times 64$.
3. Max pool'ingo lygis – langelio dydis 2×2 , įeinančio paveikslėlio dimensijos $64 \times 16 \times 128$, išeinančio paveikslėlio dimensijos $32 \times 8 \times 128$.

- 2 pilnai sujungti lygiai:

1. Pilnai sujungtas lygis – įeinančio paveikslėlio dimensijos $32 \times 8 \times 128$, išeinančių signalų kiekis – 2048.
2. Pilnai sujungtas lygis – įeinančių signalų kiekis – 2048, išeinančių signalų kiekis – 1.



6 pav. Neuroninio tinklo architektūra

2.2. Numerio simbolių atpažinimui skirtas neuroninis tinklas

Tesseract programa nuo 4.00 versijos integravo naują neuroninio tinklo pagrindu veikiančią teksto eilučių atpažinimo posistemę. Pirminis idėjos šaltinis kilo iš *OCROPUS* sistemos, kuri panaudodama Python programavimo kalbą įgyvendino LSTM veikimą. Tačiau tai buvo visiškai perdaryta panaudojus C++ kalbos ypatumus. Neuroninio tinklo sistema Tesseract programoje egzistuoja jau nuo *TensorFlow* atsiradimo ir taip pat su ja yra suderinama, kadangi naudojama tos pačios sintaksės neuroninio tinklo modelio aprašymo kalbą (VGSL).

Pagrindinė VGSL idėja yra, kad nebūtina išmokti daug naujų dalykų, kad būtų įmanoma sukurti ir apkomyti neuroninį tinklą. Nereikia mokytis *Python* programavimo kalbos, *TensorFlow* bibliotekos ar net rašyti C++ programinio kodo. Užtenka įvaldyti VGSL kalbos sintaksines ypatybes, kad būtų įmanoma taisyklingai sudaryti neuroninį tinklą.

2.2.1. Bendrai apie LSTM

LSTM yra

LSTM tinklo paslėptame rekurentiniame lygyje yra specialūs vienetai – atminties blokai. Atminties blokuose yra atminties ląstelių saugančių laikiną tinklo būseną pridedant dar specialius dauginamuosius vienetus, taip vadinamus vartus, kurie kontroliuoja informacijos srauto tėkmę. Kiekvienas atminties blokas originalioje architektūroje yra sudarytas iš įvesties ir išvesties vartų. Įvesties vartai kontroliuoja įvesties aktyvacijų informacijos tėkmę į atminties ląstelę. Išvesties vartai kontroliuoja ląstelių aktyvacijos išvesties informacijos tėkmę atgal į tinklą. Vėliau į atminties bloką pridedami atminties praradimo vartai. Tai sprendžia vieną iš LSTM modelio silpnųjų, kai įvestis nėra suskirstyta į atskiras sekas ir apdorojimas niekad nesibaigia. Atminties praradimo vartai keičia vidinę ląstelės būseną prieš patalpinant ją į ląstelę per rekurentinį ryšį su pačia savim, tuo pačiu įgalinant pamiršti ar ištrinti ląstelės atmintį. LSTM tinklas skaičiuoja sąryšius nuo įvesties sekos

$$x = (x_1, \dots, x_T)$$

iki išvesties sekos

$$y = (y_1, \dots, y_T)$$

atlikdamas tinklo vienetų aktyvacijų funkcijas pagal žemiau nurodytas formules iš eilės nuo $t = 1$ iki T :

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i)(1)$$

$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f)(2)$$

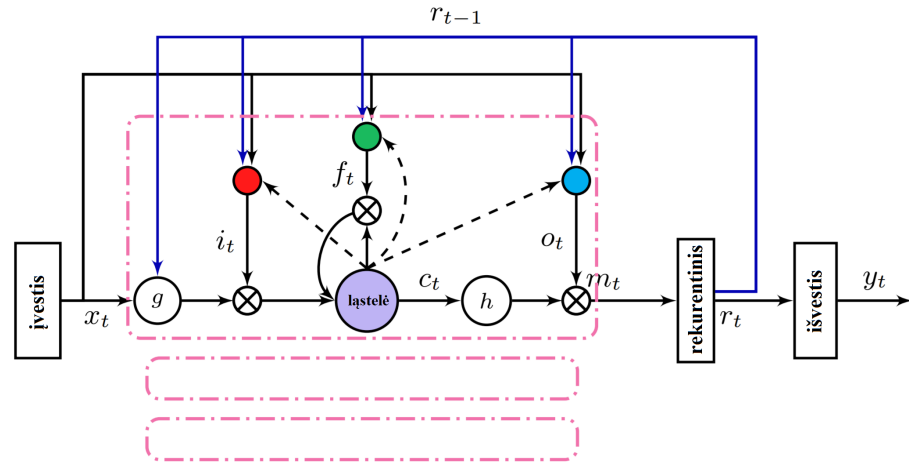
$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c)(3)$$

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o)(4)$$

$$m_t = o_t \odot h(c_t)(5)$$

$$y_t = \phi(W_{ym}m_t + b_y)(6)$$

kur W apibrėžia svorių matricas (pvz.: W_{ix} yra svorių matrica nuo įvesties vartų iki pačios įvesties), W_{ic} , W_{fc} , W_{oc} yra įstrižos svorių matricos skirtos skylių jungtims ¹, b apibūdina išvesties vektorius, kurie neturi jokios įvesties, σ yra logistinė sigmoidinė funkcija, i , f , o , c yra atitinkamai įvesties vartai, atminties praradimo vartai, išvesties vartai ir ląstelės aktyvacijos vektoriai, kurių visų dydis yra toks pat kaip ir ląstelės išvesties aktyvacijos vektoriaus m , \odot yra vektorių elementas, g ir h yra ląstelės įvestis ir ląstelės išvesties aktyvacijos funkcijos, \tanh ir ϕ yra tinklo išvesties aktyvacijos funkcija – *Softmax* [SSB14].



7 pav. LSTM atminties blokai

2.2.1.1. Rekurentinis neuroninis tinklas

Rekurentinis neuroninis tinklas – skirtingai nuo grįžtamuoju ryšiu grįstų neuroninių tinklų, yra rekursinio dirbtinio neuroninio tinklo variantas, kuriame ryšiai tarp neuronų sudaro apskritą ratą. Tai reiškia, kad išvedimo rezultatas priklauso ne tik nuo dabartinių įvesčių, bet ir nuo ankstesnio etapo neuronų būklės. Šis metodas leidžia vartotojams išspręsti problemas, susijusias su balso ar kalbos atpažinimu. Atlikti tyrimai rodo, kad įmanoma sukurti rekurentinį neuroninį tinklą, kuris gali generuoti naujus sakinius ir dokumentų santraukas.

¹Skylutės jungtis (angl. peephole connection) – LSTM tinklo ypatybė, kai vartai priklauso ne tik nuo buvusios paslėptos būsenos, bet ir nuo buvusios vidinės būsenos)

2.2.2. Integracija su Tesseract

Integruota neuroninio tinklo posistemė gali būti panaudojama kaip papildinys esamai analizės sistemai atpažįstant tekstą dideliame dokumente arba gali būti naudojama kartu su išoriniu teksto detektoriumi, kad atpažintų tekstą iš vienos teksto eilutės atvaizdo.

Nuo 4.00 versijos neuronio tinklo pagrindu veikiantis atpažinimo būdas Tesseract programoje yra numatytasis.

2.2.3. Sisteminiai reikalavimai

Nauja programos versija naudoja iki 10 kartų daugiau kompiuterio procesoriaus resursų nei senesnės Tesseract versijos, tačiau jei naudojamas kompiuteris ir platforma palaiko žemiau aprašytas funkcijas, resursų naudojimas gali sumažėti:

- *OpenMP* leidžia naudoti iki 4 procesoriaus branduolių vienu metu, jei juos procesorius turi.
- *Intel/AMD* procesoriai, kurie palaiko *SSE* ir/ar *AVX* technologiją, turi pranašumą naudojant *SIMD* branduolio matricų daugybos operacijų išlygiagretinimą.
- Kompiuteryje, kuris turi bent 4 branduolius, *AVX*, nesudėtingą anglų kalbos tekstą paveikslėlyje, atpažinimas užtrunka dvigubai ilgiau bei naudoja 7 kartus daugiau procesoriaus resursų nei ankstesnės versijos, nors Hindi kalbos atpažinimas trunka netgi greičiau nei senesnėse versijose bei naudoja tik nežymiai daugiau procesoriaus resursų.

Jei šių paminėtų komponentų nėra sistemoje, egzistuoja lėtesnė C++ kalbos implementacija, kuri vis dėlto sugeba atlikti paskirtą darbą.

2.2.4. Įgyvendinimo pagrindai

Visi neuroninio tinklo lygių tipai yra paveldėti iš bazinės *Network* klasės. *Plumbing* subklasė yra bazinė kitų tinklo lygių, kurie įvairiomis operacijomis (grupuojant keletą lygių; keičiant įvestį ir išvestį) manipuliuoja kitais lygiais, klasė.

2.2.5. Naujo tinklo lygio pridėjimas

Naujas tinklo lygis turi būti paveldimas iš klasės *Network* ar *Plumbing* ir įgyvendinti bent vieną virtualų metodą:

- *spec*, kuris grąžina *String* tipo eilutę, kuri buvo naudojama sukurti šiam tinklo lygiui.

- *Serialize/DeSerialize* – skirtas išsaugoti/atkurti tinklo lygį iš/į failą.
- *Forward* – skirtas treniravimo metu vykdyti tinklo lygį nurodant kryptį į priekį.
- *Backward* – skirtas treniravimo metu vykdyti tinklo lygį nurodant kryptį atgal.

Lygiai, kurie turi svorius taip pat turi įgyvendinti *Update* metodą, kuris atnaušina svorius naudodamas rinkinį nuolydžių. Taip pat yra keletas kitų metodų, kurie turėtų būti įgyvendinti, priklausomai nuo specifinių tinklo lygio reikalavimų:

- *NetworkBuilder* klasė turi būti pakeista, kad būtų galima apdoroti naujo tipo specifikaciją.
- *NetworkType* klasifikatorius turi būti papildytas nauju tipu.
- Naujo tipo atitinkamas įrašas turi būti pridėtas į lauką *Network::kTypeNames*.
- *Network::CreateFromFile* metodas turi būti modifikuotas, kad galėtų būti deserializuotas naujo tinklo lygio tipas.
- Kaip ir su kiekvienu nauju kodu, *lstm/Makefile.am* failas turi būti papildytas naujais failų pavadinimais.

2.2.6. VGSL specifikacijos

Kintamo dydžio grafų aprašymo kalba (angl. Variable-size Graph Specification Language) įgalina lengvai aprašyti neuroninį tinklą, susidarantį iš konvoliucijų ar LSTM tinklo ypatybių, kuris gali apdoroti kintamo dydžio paveikslėlius panaudojant vienos teksto eilutės ilgio aprašytą tinklo specifikaciją.

VGSL pritaikymas

VGSL kalba sukurta aprašyti neuroniniams tinklams, kurie:

- Kintamo dydžio (tinka ir fiksuoto dydžio) paveikslėlius naudoja kaip įvestį (vienoje ar dvejose dimensijose).
- Gauna rezultatą kaip reikšmių matricą, tekstą ar kategoriją.
- Konvoliucijos ir LSTM tinklai yra pagrindinis skaičiavimo komponentas.

Modelį aprašančios teksto eilutės įvestis ir išvestis Neuroninio tinklo modelį aprašo teksto eilutė, kurioje yra aprašomos įvesties, išvesties ir tinklo lygių specifikacijos. Pavyzdys:

[1,0,0,3 Ct5,5,16 Mp3,3 Lfys64 Lfx128 Lrx128 Lfx256 01c105]

Pirmi 4 numeriai aprašo įvesties dydį ir tipą. Tai atitinka TensorFlow sistemos paveikslėlio tenzoriaus konvenciją: [paketas, aukštis, plotis, gylis]. Šiuo atveju paketas yra ignoruojamas, bet gali būti panaudotas aprašant treniravimo paketo dydį. Aukštis ir/ar plotis gali būti lygus 0, tokiu būdu jie tampa kintamo dydžio. Nenulinės aukščio ir/ar pločio reikšmės reiškia, kad visi įvesties paveikslėliai bus vienodo dydžio arba bus suspausti iki galimo dydžio jei reikės. Gylio reikšmė 1 nurodo, kad paveikslėlis yra juodai baltas, reikšmė 3 nurodo, kad naudojamos visos spalvos. Yra specialus atvejis, kai nurodomas gylis su kitokia reikšme nei 1 ar 3 ir aukščiu – 1. Tokiu atveju tai bus traktuojama kaip vertikalių pikselių juostų seka. Paskutinis žodis nurodo apibūdina išvestį:

- Bendrinis išvesties formatas su n klasių – $O(2|1|0)(l|s|c)n$:
 - 2 (reikšmių matrica) – išvestis yra dviejų dimensių įvesties vektorių žemėlapis.
 - 1 (seka) – išvestis yra vienos dimensijos vektoriaus reikšmių seka.
 - 0 (kategorija) – išvestis yra vieno vektoriaus reikšmė.
 - l naudoja logistinę netiesinę funkciją, įgalinant išvesti keletą rezultatų bet kuriai išvesties vektoriaus reikšmei.
 - s naudoja Softmax netiesinę aktyvacijos funkciją, išvedant vieną rezultatą kiekvienai reikšmei.
 - c naudoja Softmax su CTC aktyvacijos funkciją. Gali būti naudojama tik su seka.
- Klasių skaičius yra ignoruojamas (palikta dėl suderinamumo su TensorFlow) ir tikras skaičius paimamas iš *unicharset* failo.

2.2.7. Vidinių tinklo lygių sintaksė

Žemiau aprašomos funkcinės, *plumbing* operacijos bei pateikiami jų pavyzdžiai.

Funkcinės operacijos Egzistuoja 5 skirtingos funkcinės operacijos:

- $C(s|t|r|l|m)<y>,<x>,<d>$ – vykdoma konvoliucija naudojant y , x langelį, nenaudojant sutraukimo, su atsitiktiniu užpildu, d išvestimi bei $s|t|r|l|m$ aktyvavimo funkcija.

- $F(s|t|r|l|m)<d>$ - pilnai jungus lygis su $s|t|r|l|m$ aktyvavimo funkcija ir d išvestimi. Sumažina aukštį ir plotį iki 1. Susijungia su kiekviena įvesties y , x bei gylio pozicija, sumažindamas aukštį, plotį iki 1 ir sugeneruodamas $<d>$ vektorių kaip išvestį. Įvesties aukštis ir plotis turi būti konstantos.
- $L(f|r|b)(x|y)[s]<n>$ - LSTM ląstelė su n išvesčių:
 - f - leidžia tik į priekį judantį LSTM lygį.
 - r - leidžia tik priešinga kryptimi judantį LSTM lygį.
 - b - leidžia abiejomis kryptimis judantį LSTM lygį.
 - Operacija veiks tik su x arba y kryptimi, ignoruojant kitą kryptį.
 - s - neprivalomas argumentas, kuris grąžina kaip rezultatą tik paskutinį žingsnį, su-
traukdamas dimensiją iki vieno elemento.
- $LS<n>$ - tik į priekį x kryptimi judanti LSTM ląstelė su integruota *Softmax* aktyvacijos funkcija.
- $LE<n>$ - tik į priekį x kryptimi judanti LSTM ląstelė su integruota *Softmax* aktyvacijos funkcija ir binariniu atkodavimu.

Aukščiau paminėtos raidės ($s|t|r|l|m$) reiškia vieną iš aktyvacijos funkcijų:

- s - sigmoido funkcija.
- t - hiperbolinio tangento funkcija.
- r - *Relu* funkcija.
- l - linijinė funkcija.
- m - *Softmax* funkcija.

Pavyzdžiai:

- $Cr5,5,32$ - 5×5 Relu konvoliucija su 32 filtrais.
- $Lfx128$ - tik į priekį judantis LSTM lygis, x dimensijoje turintis 128 išvestis, laikydamas y dimensiją nepriklausoma.

- L_{fys64} – tik į priekį judantis LSTM lygis, y dimensijoje turintis 64 išvestis, laikydamas x dimensiją nepriklausoma ir sutraukdamas y dimensiją iki 1 elemento.

Plumbing ops *Plumbing* operacijos leidžia konstruoti pakankamai kompleksiškus grafus:

- [...] – Vykdyti ... neuroninius tinklus nuosekliai lygiais.
- (...) – Vykdyti ... neuroninius tinklus lygiagrečiai, jungiant jų išvestis į gylį.
- $S_{\langle y \rangle, \langle x \rangle}$ – Pakeisti dviejų dimensijų įvestį susitraukimo koeficientu y,x, sutvarkant duomenis padidinant įvesties gylį koeficientu xy.
- $Mp_{\langle y \rangle, \langle x \rangle}$ – pritaikyti *Maxpool* operaciją kiekvienam stačiakampiui (y, x), gaunant vienintelę reikšmę.

Pavyzdys: Vienos dimensijos LSTM tinklas, galintis tiksliai atpažinti tekstą

[1,1,0,48 Lbx256 01c105]

Lygių aprašymas (įvesties lygis apačioje, išvesties lygis viršuje):

- 01c105: Išvesties lygis, pagaminantis vienos dimensijos seką, treniruotą su CTC, išvedantis 105 klases.
- Lbx256: Dvikryptis LSTM lygis judantis x kryptimi su 256 išvestimis.
- 1,1,0,48: Įvestis yra juodai baltas paveikslėlis, kurio aukštis yra 48 pikseliai, laikomas kaip vienos dimensijos vertikalinių pikselių seka.
- []: Tinklas visada vykdo lygius nuosekliai.

Šis sukurtas tinklas gerai veikia atpažįstant tekstą, tol kol įvesties paveikslėlis normalizuotas vertikalioje padėtyje.

Pavyzdys: Keletos lygių LSTM tinklas, galintis tiksliai atpažinti tekstą

[1,0,0,1 Ct5,5,16 Mp3,3 Lfys64 Lfx128 Lrx128 Lfx256 01c105]

Lygių aprašymas (įvesties lygis apačioje, išvesties lygis viršuje):

- 01c105: Išvesties lygis, pagaminantis vienos dimensijos seką, treniruotą su CTC, išvedantis 105 klases.

- Lfx256: Tik pirmyn judantis LSTM lygis x kryptimi su 256 išvestimis.
- Lrx128: Tik priešingai judantis LSTM lygis x kryptimi su 128 išvestimis.
- Lfx128: Tik pirmyn judantis LSTM lygis x kryptimi su 128 išvestimis.
- Lfys64: Dimensiją apibendrinantis LSTM lygis, apibendrinantis y dimensiją su 64 išvestimis.
- Mp3,3: 3x3 *Maxpool* operacija.
- Ct5,5,16: 5x5 konvoliucija su 16 išvesčių ir hiperbolinio tangento aktyvacijos funkcija.
- 1,0,0,1: Įvestis yra juodai baltas paveikslėlis.
- []: Tinklas visada vykdo lygius nuosekliai.

Šis sukurtas LSTM tinklas yra atsparesnis vertikaliesiems teksto nuokrypiams.

2.2.8. Kintamo dydžio įvestis ir apibendrinantis LSTM lygis

Kol kas vienintelis būdas sumažinti nežinomo dydžio dimensiją iki žinomo dydžio (1) yra naudojant apibendrinantį LSTM lygį. Vienas apibendrinantis LSTM lygis sumažins vieną dimensiją (x arba y), palikdamas vienos dimensijos seką. Tada vienos dimensijos seka gali būti sumažinta iki *Softmax* ar logistinės aktyvacijos funkcijos išvesties.

Toliau norint atpažinti tekstą, įvesties paveikslėlių aukštis turi būti fiksuotas arba pakeistas jų vertikalus dydis (panaudojant *Mp* ar *S* funkcijas) iki 1, arba leidžiant kintamo aukščio paveikslėlius, apibendrinantis LSTM lygis turi sumažinti vertikalią dimensiją iki vienintelės reikšmės. Apibendrinantis LSTM lygis taip pat gali būti naudojamas su fiksuoto aukščio įvestimis.

2.2.9. Modelis

Šiam konkrečiam uždaviniui, kuris turi atpažinti automobilio numerio simbolius, pasirinktas tokios konfigūracijos neuroninis LSTM tinklas:

```
[1,36,0,1 Ct3,3,16 Mp3,3 Lfys48 Lfx96 Lrx96 Lfx256 01c`head -n1 data/unicharset`]
```

Lygių aprašymas (įvesties lygis apačioje, išvesties lygis viršuje):

- 01c`head -n1 data/unicharset`: Išvesties lygis, pagaminantis vienos dimensijos seką, treniruotą su CTC, išvedantis n klasių nurodytą faile esančiame *data/unicharset*.

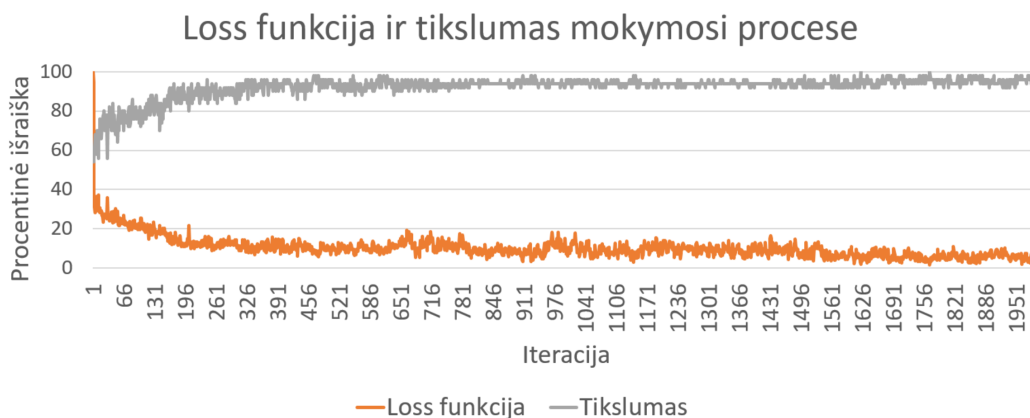
- Lfx256: Tik pirmyn judantis LSTM lygis x kryptimi su 256 išvestimis.
- Lrx96: Tik priešingai judantis LSTM lygis x kryptimi su 96 išvestimis.
- Lfx96: Tik pirmyn judantis LSTM lygis x kryptimi su 96 išvestimis.
- Lfys48: Lfys64: Dimensiją apibendrinantis LSTM lygis, apibendrinantis y dimensiją su 48 išvestimis.
- Mp3,3: 3x3 *Maxpool* operacija.
- Ct3,3,16: 3x3 konvoliucija su 16 išvesčių ir hiperbolinio tangento aktyvacijos funkcija.
- 1,36,0,1: Įvestis yra juodai baltas paveikslėlis, kurio aukštis 36 pikseliai.

3. Neuroninių tinklų apmokymas

Šiame skyriuje aprašomi veiksmai skirti apmokyti neuronius tinklus panaudojant sugeneruotus paveikslėlius.

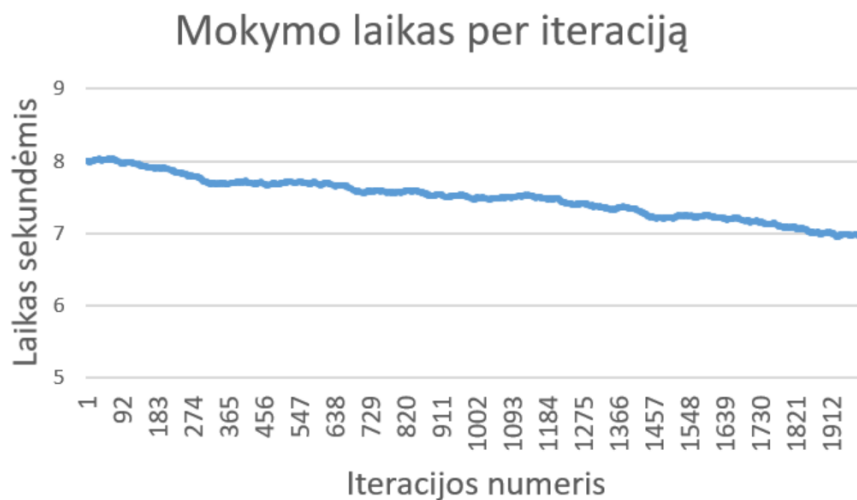
3.1. Konvoliucinio neuroninio tinklo mokymas

Apmokymui buvo naudoti 100.000 paveikslėlių, iš kurių 75.000 sudarė mokymo duomenys, o 25.000 testavimo duomenys. Vienoje iteracijoje buvo apmokoma po 50 paveikslėlių. Kas 20 iteracijų išvedami statistiniai duomenys. Kaip matome 8 paveikslėlyje, mokymosi proceso metu tikslumas priartėjo prie 100% bei pasiekė vidutinį 98% tikslumą mokymo pabaigoje.



8 pav. Loss funkcijos ir tikslumo statistika mokymosi procese

Apmokymas buvo vykdomas su ASUS GeForce GTX 1070 8GB vaizdo plokšte. Apmokyti 100.000 paveikslėlių truko apytiksliai 4h. Vidutiniškai viena iteracija truko apytiksliai 7.2s (9 pav.). Paveikslėlių generavimas buvo vykdomas tuo pačiu metu naudojantis CPU.



9 pav. Neuroninio tinklo mokymo greitis

3.2. LSTM rekurentinio neuroninio tinklo mokymas

Tesseract 4.00 versijoje pridėtas naujas atpažinimo variklis, kuris remiasi LSTM tipo rekurentiniu neuroniniu tinklu. Lyginant su ankstesnėmis versijomis, ženkliai padidėjo dokumentų tipo nuotraukų teksto atpažinimas, tačiau tai reikalauja ženkliai didesnių kompiuterio skaičiavimo resursų. Atpažįstant sudėtingas kalbas, yra didelė tikimybė, kad atpažinimas truks greičiau nei bazinė pirminė Tesseract versija.

Naudojant neuroninius tinklus teksto atpažinimui yra reikalinga žymiai daugiau duomenų modelio treniravimui, taip pat pats treniravimas trunka ilgiau nei pirminėje Tesseract versijoje. Visoms lotynų rašmenimis pagrįstoms kalboms treniravimas vyko naudojant daugiau nei 400.000 teksto eilučių bei apie 4.500 skirtingų šriftų. Su nauja versija ženkliai išaugo mokymosi laikas. Jei su ankstesne versija mokymas trukdavo nuo kelių minučių iki kelių valandų, tai su nauja 4.00 versija tai gali trukti nuo kelių dienų iki kelių savaičių. Tačiau ne visais atvejais yra naudinga treniruoti modelį nuo pradžių, priklausomai nuo situacijos, kartais užtenka pertreniruoti egzistuojantį modelį.

Išskiriami trys pagrindiniai modelio apmokymo principai:

- Esamo modelio patobulinimas. Naudojant egzistuojantį pasirinktos kalbos modelį, papildomai apmokomas su papildomais specifiniais duomenimis. Tai gali išspręsti problemas, kai norimas rezultatas nedaug skiriasi nuo jau apmokyto modelio, pvz.: truputį nestandartinis šriftas. Gali veikti su sąlyginai mažu naujų duomenų kiekiu.
- Nuimti viršutinį (ar keletą daugiau) modelio sluoksnių ir pertreniruoti naujus sluoksnius su naujais duomenimis. Jei esamo modelio patobulinimas nesprendžia esamos problemos, šis būdas dažniausiai būna kitas pasirinkimas. Viršutinio sluoksnio permokymas vis dar gali veikti treniruojant visiškai naują kalbą, tačiau tos kalbos turi būti labai panašios, kad būtų pasiektas norimas efektas.
- Apmokymas nuo nulio. Tai gali būti labai sunki užduotis, jei nėra pakankamai daug reprezentatyvių duomenų spręsti konkrečiai problemai. Jei duomenų nėra pakankamai daug, galiausiai tinklas bus permokytas, kuris puikiai susidoros tik su mokymo duomenimis, tačiau visiškai neatliks savo užduoties, kai bus paduodami realūs duomenys. Nors mokymas atrodo skiriasi, matys treniravimo žingsniai yra beveik identiški aukščiau aprašytiems, taigi tai yra visai paprasta išbandyti, atsižvelgiant į turimų duomenų bei kompiuterio resursų kiekį.

3.2.1. Atpažinimo kokybės gerinimas

Egzistuoja įvairiausių priežasčių, kodėl Tesseract atpažinimo programa nesugeba atpažinti jai paduoto teksto. Svarbu pabrėžti, kad Tesseract modelio permokymas retai padės, nebent naudojamas labai nestandartinis šriftas arba nauja dar netreniruota ir neapmokyta kalba.

3.2.1.1. Paveikslėlio apdorojimas

Pati Tesseract sistema savyje atlieka įvairius paveikslėlių apdorojimo veiksmus, pasinaudojant Leptonica biblioteka, prieš pradedant pati teksto atpažinimą. Dažniausiai Tesseract puikiai susitvarko su šita užduotimi, tačiau neišvengiamai atsiranda situacijų, su kuriomis automatiškai susidoroti nepavyksta ir dėl to pastebimai nukenčia atpažinimo tikslumas.

Jei norima pamatyti, kaip Tesseract apdorojo paveiksluką, tai galima atlikti pakeitus konfigūracinio parametro *tessedit_write_images* reikšmę į **true** kai yra leidžiama Tesseract programa. Jei paruoštas tinklo apmokymui paveikslėlis atrodo problematiškai, neišvengiamai reikės pritaikyti vieną ar daugiau paveikslėlių apdorojimo technikų prieš siunčiant apdorojimui.

Spalvų inversija

Nors senesnės Tesseract versijos (≤ 3.05) palaikė šviesų tekstą ant juodo fono be jokių problemų, nuo 4.00 versijos būtina sąlyga, kad tekstas būtų juodas, o fonas šviesus. Tam tikslui atlikti užtenka vienos komandos:

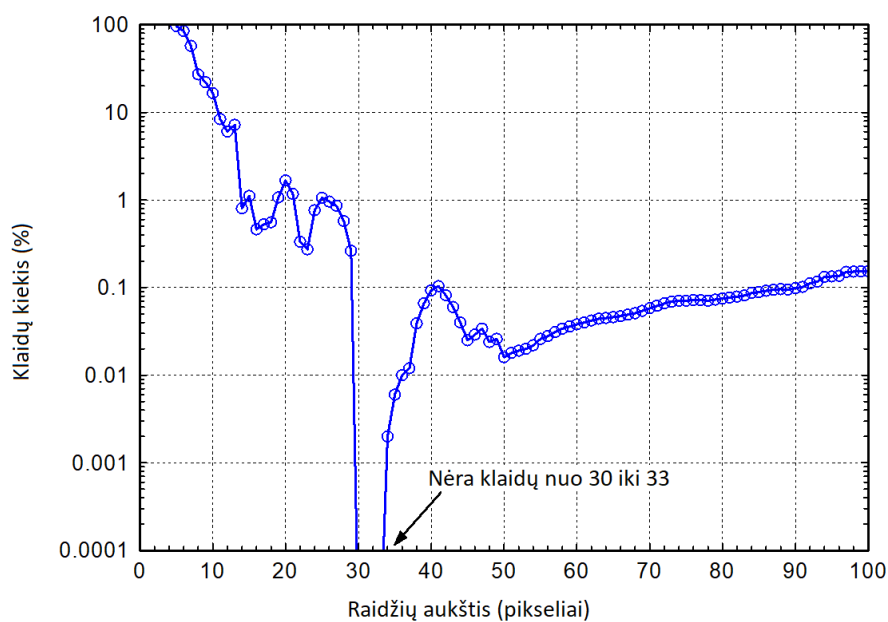
```
numpy.invert(image)
```

Dydžio keitimas

Tesseract programa geriausiai veikia, kai paduodamų paveikslėlių taškų viename colyje (angl. DPI) dydis yra bent 300, todėl labai svarbu užtikrinti, kad dydis nebūtų mažesnis.

Atliktas eksperimentas² (žiūrėti 10 pav.) parodė, kad egzistuoja optimalus raidžių aukštis, kuriam esant klaidų tikimybė mažėja iki 0. Raidžių aukščiui esant tarp 30 ir 33 pikselių, klaidų tikimybė visiškai sumažėja, todėl galima daryti prielaidą, kad labai svarbu pasirinkti tinkamą šrifto dydį ruošiant mokymo duomenis, norint pasiekti geriausių rezultatų.

²Willus Dotkom vartotojo atliktas eksperimentas. Šaltinis https://groups.google.com/forum/#!msg/tesseract-ocr/Wdh_JJwnw94/24JHDYQbBQAJ



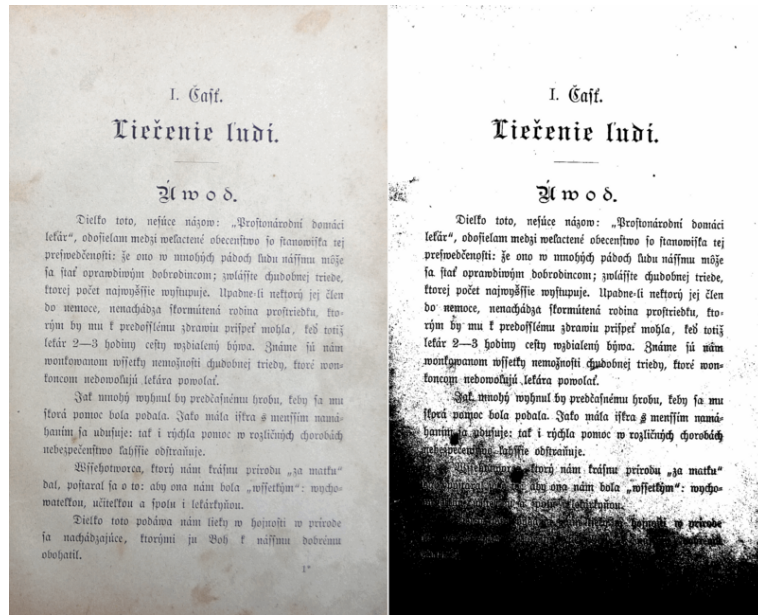
10 pav. Klaidų kiekio priklausomybė nuo raidžių aukščio

Binarizacija

Binarizacija – tai paveiksluko spalvų keitimas į juodą ir baltą. Tesseract jau turi integruotą funkcionalumą atlikti šiai užduočiai (naudojamas *Otsu* algoritmas), tačiau ne visada rezultatas gaunasi optimalus. Tai dažniausiai lemia netolygus fono tamsumas.

Jei nepavyksta išgauti geresnės kokybės nuotraukos, kuriame fono spalva būtų tolygi, yra alternatyvių ribinių verčių nustatymo algoritmų, kuriuos vertėtų išbandyti:

- *ImageJ* automatinis ribinių verčių nustatymo algoritmas (JAVA programavimo kalba).
- *OpenCV* ribinių verčių nustatymo algoritmas (Python programavimo kalba).
- *scikit-image* ribinių verčių nustatymo algoritmas (Python programavimo kalba).



11 pav. Binarizacijos algoritmo taikymo rezultatas

Triukšmo pašalinimas

Triukšmas – tai atsitiktinis netolygaus ryškumo išsibarstymas paveikslėlyje, kuris gali padaryti tekstą sunkiai ar visai neįskaitomą. Yra specifiniai triukšmo tipai, kurių Tesseract nesugeba pašalinti vykdydama binarizacijos etapą, todėl ženkliai sumažėja atpažinimo tikslumas.

- θεῶν τὸν πλάνον διήλεγεν; ἀναφανδὸν γὰρ τοὺτους ἐφησεν
ὁ τῆς ἀληθείας ἀντίπαλος μῆτε θεοὺς μῆτε ἀγαθοὺς δαι-
μονας εἶναι, ἀλλὰ τοῦ ψεύδους διδασκάλους καὶ πονηρίας
70 πατέρας. τοὺτους ὁ Πλάτων ἐν τῷ Τιμαίῳ οὐδὲ φῦσει
ἀθανάτους φησὶν. τὸν γὰρ ποιητὴν εἰρηκέναι πρὸς αὐτοὺς
λέγει· „ἀθάνατοι μὲν οὐκ ἐστὶ οὐδ’ ἄλλοι τὸ πᾶμπαν
οὐτι μὲν δὴ λυθήσεσθε, τῆς ἐμῆς βουλήσεως τυγχόντες.“
καίτοι γε Ὀμήρῳ τάναντία δοκεῖ ἀθανάτους γὰρ αὐτοὺς
πανταχῇ προσονομάζει· „οὐ γὰρ σίτον“ φησὶν „ἔδουσ’ οὐ
πίνονσ’ αἰδοπα οἶνον· τούνεκ’ ἀναίμονες εἰσι καὶ ἀθάνατοι 10
καλέονται.“
- 71 Τῶσαντ’ ἡ παρὰ τοῖς ποιηταῖς καὶ φιλοσοφοῖς περὶ τῶν
οὐκ ὄντων μὲν, καλουμένων δὲ θεῶν διαμάχη. τοῖς τοῖς καὶ
νεῶς ἐδομήσαντο καὶ βωμοὺς προσωκοδόμησαν καὶ θυσίαις
ἐτίμησαν καὶ εἶδη τινὰ καὶ εἰκασμάτα ἐκ ξύλων καὶ λίθων 15
καὶ τῶν ἄλλων ὕλων διαγλύψαντες, θεοὺς προσηγόρευσαν
τὰ χειρόμνητα εἰδῶλα καὶ τὰ τῆς Φειδίου καὶ Πολυκλείτου
καὶ Πραξιτέλους τέγνης ἀγάλματα τῆς θείας προσηγορίας
72 ἡξίωσαν. τοῦτον δὲ τοῦ πλάνου κατηγοροῦν Ξενοφάνης ὁ
Κολοφώνιος τοιαῦτα φησὶν· „ἀλλ’ οἱ βροτοὶ δοκοῦσι γεννᾶ- 20
σθαι θεοὺς καὶ ἴσῃν τ’ αἰσθῆσιν ἔχειν φωνὴν τε δέμας τε.“
καὶ πάλιν· „ἀλλ’ εἴ τοι χεῖρας εἶχον βόες ἢ λέοντες ἢ
γράψαι χεῖρεσσι καὶ ἔργα τελεῖν ἄπερ ἄνδρες, ἵπποι μὲν θ’
ἵπποισι, βόες δὲ τε βανσίῳ ὁμοίας καὶ θεῶν ἰδέας ἔγραφον
καὶ σώματ’ ἐποίουν τοιαῦθ’, οἷόν περ καὶ τοὶ δέμας εἶχον 25

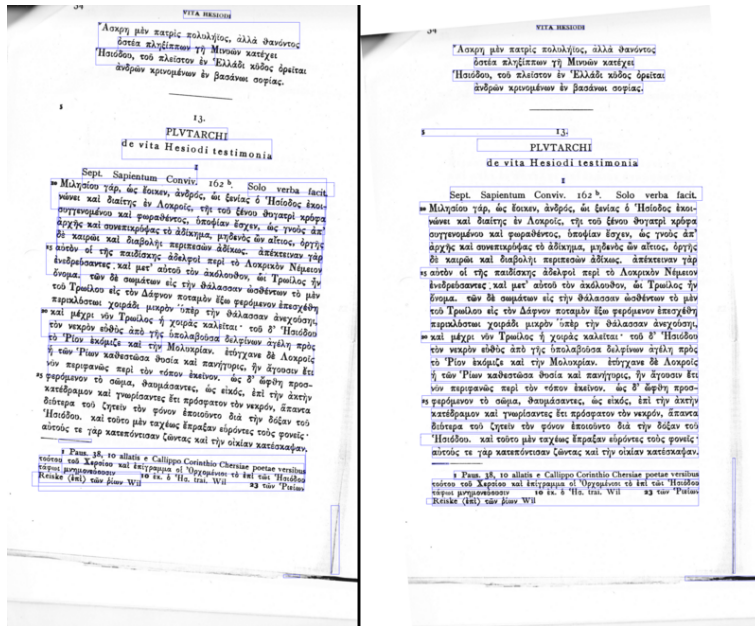
6—7: Eus. Pr. XI 32, 4. XIII 18, 10 (Plat. Tim. p. 41 B).
9—11: Hom. E. 341—342. || 19.—p. 89, 1: Clem. Str. V 14, 100
= Eus. Pr. XIII 13, 38 (Xenophan. fr. 14—15)

1 ἐφησεν: ἐδήσαν in ἐδήλασεν corr. S. | 7 οὐτι: οἱ BLS:
ὁ τε V | λυθήσεσθαι M. corr. Mgr.: λυπηθήσεσθαι. L1 | 8 γε om.
BLMCV | 9 πανταχᾶ K: πανταχοῦ BL | ἔδουσιν codd. | οὐ
(posteriore loco): οὐδὲ BLMCV. | 10 πίνοισιν codd. | 13 περ:
παρὰ V | 14 νεῶς M. | ἐδομήσαντο BS: ἐδόμησαν K | καὶ θυσίαις
ἐτίμησαν om. S, sed posuit infra, post λίθων | 15 εἶδη BL:
ἔδη K | 17 χειρόμνητα MCV | 20 τοιαῦτα BL | βροτοί M | 21 α’
αἰσθῆσιν: ταῖς τιθήσιν K | 22 εἰ: ἢ L. e. corr. | τοι: τῇ V | ἔχον
K | ἢ λέοντες ἢ ἐλέφαντες MSCV | 23 χεῖρεσσι MS | ἄπαν M¹ |
θ’: μεθ’ MSC | 24 δὲ om. V | ἰδέας BLSO, sed corr. S

12 pav. Pašalintas triukšmas

Pasukimas / Iškreipimas

Iškreiptas paveikslėlis būna tada, kai yra nuskanuojamas lapas kreivei. Tesseract linijų atpaži-
nimo tikslumas sumažėja jei puslapis nėra visiškai horizontalus, o tai įtakoja patį teksto atpažinimą.
Norint išspręsti šią problemą, reikia pakreipti puslapį taip, kad tekslo linijos būtų horizontalios.

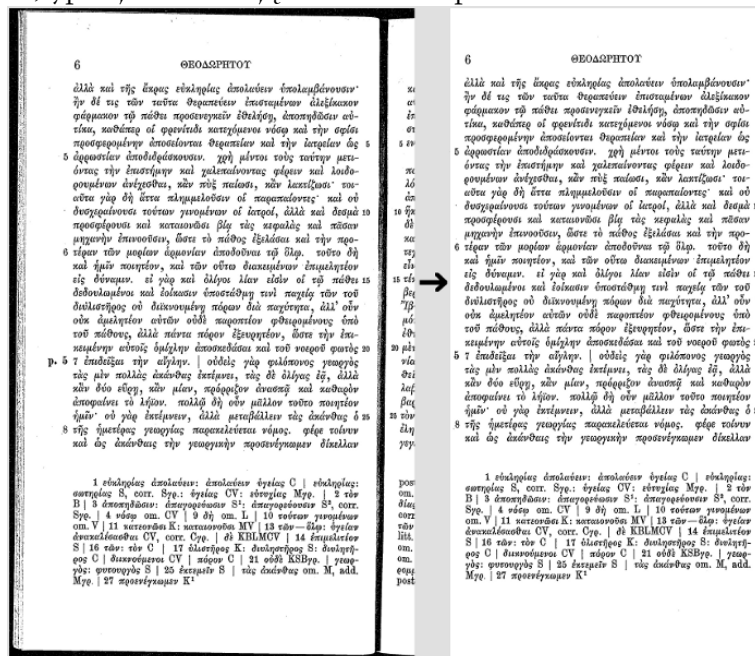


13 pav. Iškreipto puslapio išlyginimas

Kraštinės

Skanuotų puslapių kraštinių naikinimas

Skanuoti puslapiai dažnai turi tamsias kraštines aplinkui tekstą. Tai dažnai gali būti atpažįstami kaip papildomi simboliai, ypač jei skiriasi jų formos ir atspalviai.



14 pav. Puslapio kraštinių naikinimas

Tekstas be kraštinių

Jei norimas atpažinti tekstas visiškai neturių kraštinių ir yra nuo krašto

iki krašto, Tesseract programa gali turėti sunkumų bandant atpažinti tekstą. Panaudojant vieną komandą, lengvai galima pridėti kraštines iš visų pusių (naudojama *ImageMagick®* programa):

```
convert input.jpg -bordercolor White -border 10x10 output.jpg
```

Permatomumas / alfa kanalas

Kai kurie paveikslėlių formatai (pvz.: png) turi alfa kanalą, kuris suteikia galimybę saugoti permatomumo reikšmę nuotraukoje. Alfa kanalu dažniausiai nusakomas paveikslėlio skaidrumas. Paprastai prie 24 nuotraukos bitų, kuriuose kiekvienai iš trijų pagrindinių spalvų skiriama po 8 bitus, pridedami papildomi 8 bitai, kurie saugo skaidrumo informaciją.

Tesseract 3.0x versijos tikisi, kad pats vartotojas pateiks paveiksluką jau su panaikinta alfa kanalu. Tai gali būti padaroma su tokia komanda (naudojama *ImageMagick*® programa):

```
convert input.png -alpha off output.png
```

Tesseract 4.00 versijoje yra funkcionalumas, kuris pats pašalina alfa kanalą naudojant *Leptonica* programos komandą *pixRemoveAlpha()*. Ši komanda panaikina alfa kanalą suliedama jį su baltu fonu. Kartais (pvz.: filmų subtitrų atpažinimas) tai gali sukelti problemų, todėl vartotojai turėtų patys panaikinti alfa kanalą arba pritaikyti spalvų inversiją.

3.2.1.2. Puslapių skirstymo metodas

Tesseract programos standartinis veikimo principas pagrįstas tuo, kad programa tikisi paveikslėlio puslapio pavidalu su jame esančiu tekstu. Tačiau, jei norima atpažinti tik dalį teksto, yra įvairiausių teksto skirstymo parametrų, kurių reikia nurodyti naudojant komandą *--psm* ir nurodant komandos numerį.

0. Orientacija ir rašto aptikimas (OSD).
1. Automatinis puslapio skirstymas su rašto aptikimu (OSD).
2. Automatinis puslapio skirstymas, bet be rašto aptikimo (OSD) ir be simbolių atpažinimo (OCR).
3. Pilnai automatinis puslapio skirstymas, bet be rašto aptikimo (OSD) (Numatytasis režimas).
4. Vienas teksto stulpelis.
5. Vienas vertikalčiai išlygiuoto teksto blokas.

6. Vienas teksto blokas.
7. Paveikslėlį laikyti kaip vieną teksto liniją.
8. Paveikslėlį laikyti kaip vieną žodį.
9. Paveikslėlį laikyti kaip vieną žodį apskritime.
10. Paveikslėlį laikyti kaip vieną simbolį.
11. Išmėtytas tekstas. Rasti kuo daugiau teksto nesilaikant jokios tvarkos.
12. Atpažinti išmėtytą tekstą su rašto aptikimu (OSD).
13. Neapdorota eilutė. Paveikslėlį laikyti kaip vieną teksto liniją, išvengiant specifinių Tesseract gudrybių.

3.2.1.3. Žodynai, žodžių sąrašai, šablonai

Tesseract programa optimizuota taip, kad geriausiai atpažintų sakinius, susidarančius iš žodžių. Jei yra bandoma atpažinti nestandartinės struktūros tekstus (pvz.: sąskaitas, čekius, prekių sąrašus, kodus), yra keletas papildomų būdų, kaip būtų galima pagerinti atpažinimo tikslumą.

Pirmiausiai reikia įsitikinti, kad yra pasirinktas tinkamas puslapio skirstymo būdas. Tai užtikrina, kad bus efektyviausiai ieškoma teksto.

Žodynų atjungimas, kuriuos naudoja Tesseract turėtų pagerinti atpažinimą, jei dauguma teksto nėra žodyne esantys žodžiai. Norint išjungti funkcionalumą, kai naudojami Tesseract žodynai, reikia nurodyti *FALSE* reikšmę šiems konfigūraciniams parametrams: *load_system_dawg* ir *load_freq_dawg*.

Taip pat yra galimybė pačiam vartotojui prisidėti norimus žodžius į Tesseract programą, kurie padės atpažinimo varikliui geriau suprasti žodžius. Be žodžių, yra galimybė prisidėti simbolių sekų šablonus, kurie dar labiau padės pagerinti tikslumą.

3.2.2. Rinkmenų pasiruošimas

Bendrai mokymo žingsniai yra tokie:

1. Pasiruošti norimą apmokyti tekstą.
2. Sugeneruoti paveiksluką su tekstu + *box* failu.

3. Sukurti *unicharset* failą.
4. Iš *unicharset* sukurti pradinę apmokymo duomenų failą ir nebūtiną žodynų informaciją.
5. Paleisti *Tesseract*, kad apdorotų paveikslėlį ir *box* failą bei sukurtų apmokymo duomenų rinkinį.
6. Paleisti treniravimą su sukurtu duomenų rinkiniu.
7. Sujungti duomenų failus.

Norint atlikti LSTM rekurentinio tinklo mokymą *Tesseract* 4.0 versijos aplinkoje, reikia sugeneruoti atitinkamo formato mokymo rinkmenas. Kiekvienas sugeneruotas automobilio numeris turi turėti 5 skirtingus rinkmenas:

- .box formato -
- .lstmf formato -
- .gt.txt formato - tekstinė rinkmena, kurioje yra tekstas, kuris yra pavaizduotas paveikslėlyje.
- .txt formato -
- .tif formato - paveikslukas, išsaugotas TIFF formatu.

--TODO <https://github.com/tesseract-ocr/tesseract/wiki/TrainingTesseract-4.00> pridėti kaip vyko tinklo mokymas 2h Makefile 1h statistika 1h

4. Vaizdo atpažinimas

Šiame skyriuje aprašoma kaip panaudojami apmokyti neuroniniai tinklai skirti atlikti jiems paskirtas užduotis.

4.1. Numerio rėmelio atpažinimas

Norint aptikti ir atpažinti realiuose paveikslėliuose numerio rėmelį, į neuroninį tinklą paduodamos 128x64 pikselių dydžio paveikslėlio dalys, kaip jau buvo aprašyta 1.1 skyriuje. Atpažįstant realius paveikslėlius, naudojama kitokia neuroninio tinklo architektūra. Paskutiniai du lygmenys vietoj to, kad būtų pilnai sujungti, yra konvoliuciniai. Taip pat pradinio paveikslėlio dydis neturi būti 128x64 pikselių, o gali būti bet koks. Idėja tokia, kad pilno dydžio paveikslėlis gali būti paduodamas į neuroninį tinklą suskaidant jį į dalis slenkančio langelio principu, bei kiekvienai iš jų grąžinant rezultatą, ar rėmelis egzistuoja. Naudojant vienodą neuroninį tinklą visoms paveikslėlio dalims yra pranašesnis nei atskiri neuroniniai tinklai, kadangi slenkantys langai dalinsis dauguma konvoliucinių savybių tarpusavyje, todėl nereikės kiekvieną kart atlikti naujų skaičiavimo operacijų.

Slenkančio lango principu veikiančio neuroninio tinklo rezultatai:



15 pav. Slenkančio lango principu gauti rezultatai

Žali stačiakampiai (15 pav.) vaizduoja regionus kur tikimybė, kad rėmelis egzistuoja yra didesnė arba lygi 99%. Tai padaryta tokiu tikslu, kadangi mokymo duomenų aibėje apie 50% paveikslėlių yra su egzistuojančiu numerio rėmeliu, kai realiame pasaulyje paveikslėlių su numerio rėmeliais yra daug mažiau. Jeigu būtų naudojama 50% tikimybė atrinkti teisingiems paveikslė-

liams, tai būtų neapsisaugota nuo pasitaikančių panašių paveikslėlių atitikmenų.

Norint panaikinti perteklinius dublikatus, pritaikomas Non-Maximum Suppresion ³ algoritmas, kuris tarp visų besikertančių stačiakampių palieka tik didžiausią tikimybę turinčią reikšmę[GDD⁺14].

Gavus likusį vieną stačiakampį (16 pav.), pagal to objekto koordinatas iškerpamas paveikslėlis ir gaunamas toks rezultatas:



16 pav. Iškirptas gautas rezultatas

4.2. Numerio simbolių atpažinimas

--TODO programa kuri naudoja tsesseract su istreniruotu tinklu 1h paveiksliukai 1h

³Non-Maximum Suppresion angl. – ne maksimalios reikšmės slopinimo algoritmas

Rezultatai

--TODO 20min

1. Pasinaudojus paveikslėlių duomenų rinkiniu susigeneruoti 1.000.000 atsitiktinių paveikslėlių su automobilio numeriais.
2. Apmokyti kursinio darbo metu sukurtą konvoliucinį neuroninį tinklą (rėmelio atpažinimui) pateikiant sugeneruotus paveikslėlius.
3. Apmokyti Tesseract LSTM rekurentinį neuroninį tinklą (teksto atpažinimui) pateikiant sugeneruotus paveikslėlius.
4. Pasinaudojus kursinio darbo metu sukurtu ir apmokytu konvoliuciniu neuroniniu tinklu atpažinti numerio rėmelį paveikslėlyje ir gauti jo koordinatas.
5. Pagal gautas koordinatas, iškirpti rėmelį ir pasinaudojus Tesseract LSTM neuroniniu tinklu atpažinti numerį bei atvaizduoti gautus rezultatus pradiname paveikslėlyje.
6. Ištestuoti tinklą su tikrais paveikslėliais, kuriuose yra lietuviški automobilių numeriai.

Išvados

--TODO 20min Išvadų skyriuje daromi nagrinėtų problemų sprendimo metodų palyginimai, siūlomos rekomendacijos, akcentuojamos naujovės. Išvados pateikiamos sunumeruoto (gali būti hierarchinis) sąrašo pavidalu. Darbo išvados turi atitikti darbo tikslą.

LITERATŪROS SĄRAŠAS

- [BSS13] Bharat Bhushan, Simranjot Singh ir Ruchi Singla. License plate recognition system using neural networks and multithresholding technique. *International journal of computer applications*, 84(5), 2013.
- [GBI⁺13] Ian J Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud ir Vinay Shet. Multi-digit number recognition from street view imagery using deep convolutional neural networks. *Arxiv preprint arxiv:1312.6082*, 2013.
- [GDD⁺14] Ross Girshick, Jeff Donahue, Trevor Darrell ir Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the ieee conference on computer vision and pattern recognition*, 2014, p.p. 580–587.
- [LS16] Hui Li ir Chunhua Shen. Reading car license plates using deep convolutional neural networks and lstms. *Arxiv preprint arxiv:1601.05610*, 2016.
- [SHT⁺15] Fabian Stark, Caner Hazırbaş, Rudolph Triebel ir Daniel Cremers. Captcha recognition with active deep learning. *Workshop new challenges in neural computation 2015*. Citeseer, 2015, p. 94.
- [Sla99] Gregory G Slabaugh. Computing euler angles from a rotation matrix. *Retrieved on august*, 6(2000):39–63, 1999.
- [Smi07] Ray Smith. An overview of the tesseract ocr engine. *Ninth international conference on document analysis and recognition (icdar 2007)*. Tom. 2. IEEE, 2007, p.p. 629–633.
- [SSB14] Haşim Sak, Andrew Senior ir Françoise Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. *Fifteenth annual conference of the international speech communication association*, 2014.
- [XHE⁺10] Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva ir Antonio Torralba. Sun database: large-scale scene recognition from abbey to zoo. *Computer vision and pattern recognition (cvpr), 2010 ieee conference on*. IEEE, 2010, p.p. 3485–3492.

Sąvokų apibrėžimai

- Tesseract – optinė ženklų atpažinimo programa, kuri geba naudoti neuroninius tinklus atpažinimui.
- Leptonica
- Tensorflow
- OpenMP
- Tenzorius – geometrinis objektas, susidedantis iš sumos komponentų, kurios yra transformuojamos pagal tiesinius sąryšius.
- Softmax
- Relu
- Plumbing
- Maxpool

Santrumpos

--TODO

- LSTM – trumpinys angl. Long short-term memory – rekurentinio neuroninio tinklo architektūra.
- TIFF
- DPI
- PNG
- OSD
- OCR
- VGSL
- SSE

- AVX
- SIMD
- CTC -
-
-
-
-