

# Raytins LT kalbos identifikavimas (OCR)

Paulius Milmantas  
Vilniaus universitetas  
Matematikos ir informatikos fakultetas  
Vilnius, Lietuva  
paulius.milmantas@mif.stud.vu.lt

**Abstract**—Darbe buvo sprsta raytins lietuvi kalbos raidi at-painimo problema. Parayta programa gali aptikti 6 imokyta raides: A, a, B, C, P, u vidutinikai su 93utenka sukurti nauj duomen rinkin pagal tam tikr format ir itreniruoti nauj 3 element aptikimo tinkl. Kadangi programoje naudojama daug skirting tinkl ir kiekvienas j atskiria tik 3 elementus, pleiant program, nebereikia apmokyti visos sistemos i naujo, utenka tik pridti nauj tinkl, programa j automatikai aptinka ir pradeda naudoti. Naujam 3 raidi poaibiui apmokyti reikia vidutinikai 50 nuotrauk kiekvienai raidei ir norint pasiekti 93platform tai vidutinikai utrunka apie 5 minutes. Naudojant daug neuronini tinkl taip pat isprsta didelio RAM naudojimo problema: pagal esamus resursus galima apskaiiuoti kiek vienu momentu galima pakrauti tinkl ir pakrauti tik tam tikr j kiek.

## I. VADAS

Uduoties tikslas: parayti sprendim, kuris i duotos nuotraukos igaut joje pateikt Lietuvik raytin tekst.

## II. METODAI

### A. Taikyta nuostoli funkcija

Darbe buvo naudota MSE (Mean Square Error) funkcija. (1)

$$MSE = \frac{1}{n} * \sum_{i=1}^n (y_i - y_i^p)^2 \quad (1)$$

### B. Optimizavimo funkcija

Optimizavimui buvo naudota stochastinio gradientinio nuolydio (SGD) optimizavimo funkcija. (2)

$$\theta^{(\tau)} = \theta^{(\tau-1)} - \eta * \nabla_{\theta} Loss(\theta^{(\tau-1)}; (x_i, y_i)) \quad (2)$$

### C. Naudojamas tinklas

Visos abels raidis yra skirstomos poaibus po 3 raides. 1 Kiekvienam poaibiui yra sukuriamas po atskir neuronin tinkl. Taip yra lengviau atlikti tinklo treniravim ir tinklo ukrovimui galima sutaupyti RAM atminties. Norint pridti daugiau duomen, utenka tinkl apmokyti tik vienam naujam poaibiui.

Tinkl sudaro 2 paslpti sluoksniai, 1 vesties ir 1 ivesties sluoksnis. Duomenys yra 64x64 dydio pilki vaizdai, todl vesties sluoksnis yra 4096 dydio. Ivesties sluoksnis yra 3 dydio, nes visi poaibiai yra sudaryti i 3 nari. Visi sluoksniai naudoja RELU aktyvacijos funkcijas.

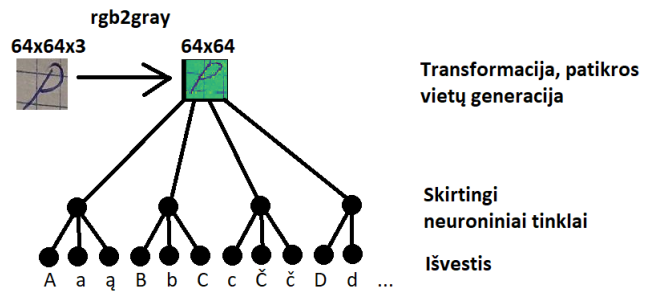


Figure 1 – Tinklo sudtis.

### D. Apdorojimo srii paieka

Srii radimui, kurias norima leisti per neuroninius tinklus, buvo naudojama objekt kratini paieka. Kadangi kiekvien piksel apibria tik vienas skaiius: pilkos spalvos stiprumas, galima eiti per paveiksliuko kiekvien piksel ir tinrinti ar jo reikm labai skiriasi nuo praeitos. Iuo metodu gaunamos visos kratins, taiau atsiranda ir daug triukmo. Triukmas paprastai aikiuose vaizduose bna nedidelis, todl j galima paalinti tikrinant kratini vientiesikum: jeigu pikselis yra aptiktas kaip kratin, tai j turi supti dar nors vienas pikselis, kuris yra laikomas kratine. Jeigu tokio pikselio alia nra, reikia tikrinamas pikselis nra kratin ir jo nereikia traukti. Rezultatas pateiktas vaizde 3.

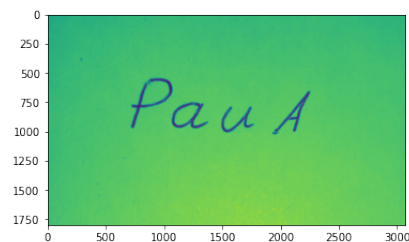


Figure 2 – Vaizdas prie apdorojim.

Radus kratines ir taikant slyg, kad vaizde nra daug paalini objekt galima aproksimuoti Y aies padt, ties kuria yra paraytas tekstas, apskaiiuojant rast kratini tak Y koordinates. Rezultatas pateiktas vaizde 4.

Radus teksto kratines yra naudojamas slenkanio lango algoritmas: einama pro paveiksliuko pikselius ir didinant imamo vaizdo plotis ir vaizdo variantai yra siuniami vaizdai pro tinklus. Jeigu didinant vaizdo plot tikimyb maja arba spjama jog yra kitas objektas, tada yra iekoma kitos raidis.

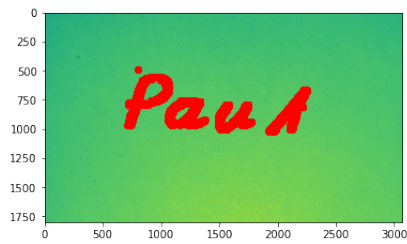


Figure 3 – Rastos kratins.

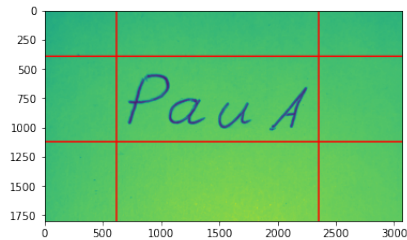


Figure 4 – Y aies radimas.

### III. DUOMENYS

Suymt duomen, kuri reikia norint imokyti model, internete nra, todl jie buvo renkami ranka. Ant lapo buvo suraomos raids ir visas lapas buvo nufotografuojamas. Gautos fotografijos buvo apdorojamos duomen ymjimo programa, kuri kaip rezultat eksportavo JSON formato fail su kiekvienos raids pozicija nuotraukoje. Pagal gaut JSON fail kiekviena raid buvo eksportuota atskir JPG fail ir atitinkamai apdorota: naudojant nearest neighbour metod sumainta iki 64x64 dydio ir panaikintas RGB kanalas.

### IV. REZULTATAI

Parayta programa gali aptikti 6 imokytyas raides: A, a, B, C, P, u vidutinikai su 93su 94,23utenka sukurti nauj duomen rinkin pagal tam tikr format ir itreniruoti nauj 3 element aptikimo tinkl. Kadangi programoje naudojama daug skirting tinkl ir kiekvienas j atskiria tik 3 elementus, pleiant program, nebereikia apmokyti visos sistemos i naujo, utenka tik pridti nauj tinkl, programa j automatikai aptinka ir pradeda naudoti. Naujam 3 raidi poaibiui apmokyti reikia vidutinikai 50 nuotrauk kiekvienai raidei ir norint pasiekti 93platform tai vidutinikai utrunka apie 5 minutes. Naudojant daug neuronini tinkl taip pat isprsta didelio RAM naudojimo problema: pagal esamus resursus galima apskaiiuoti kiek vienu momentu galima pakrauti tinkl ir pakrauti tik tam tikr j kiek.

Cituojuame altin [?].