BABES-BOLYAI UNIVERISTY CLUJ-NAPOCA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

BACHELOR THESIS

Static Sign-Language Recognition using RetinaNet

Supervisor:
Prof. Dr. Laura DIOSAN

Author:
Norbert-Cristian BERECZKI

2018

Abstract

In the last three decates there have been significant contributions related to sign-language recognition. So it is imperative to continue contributing to this topic in order to improve how deaf-mute people interact with society. Many papers have been published presenting methods using either hand-colored gloves or using special devices such as Microsoft Kinect. In the real world, most of the time, devices like these are difficult to aquire. However, image streams and videos are more common means of procuring data from a user. In this work we present a real-time American-Sign-Language translator. In order to develop such a tranlator we employed Machine Learling techniques. Our model is based on transfer learning by using a pre-trained RetinaNet convolutional network and fine-tune using University of Massey dataset and an open source dataset found on github [6].

This work is the result of my own activity. I have neither given nor received unauthorized assistance on this work.

Contents

Introduction

Absolventul va prezenta rezumativ tema tratat? relativ la enun?ul problemei, obiectivele urm?rite, rolul aplica?iei ?i structura lucrarii, precum ?i leg?tura dintre capitole.

Lucrarea de fa?? ofer? o vedere de ansamblu a ...

Capitolul 2 prezint? ...

?n capitolul 3 sunt definite no?iunile de ...

Capitolul 4 prezint? ...

?n capitolul 5 prezint? ...

Lucrarea se ?ncheie cu concluzii ?i direc?ii de cercetare.

Chapter 1

Scientific Problem

1.1 Problem definition

The problem which we want to solve is of great importance to the deaf-mute community.

In the recent couple of decades there have been many contributions related to sign-language recognition. Most of this work is concentrated on hand-colored gloves and Microsoft Kinect. However, in real life, it would be much easier for the user to have a recognition system that only requires image streams.

Now, we have reduced the problem to classifying certain images in the stream. We proposed a solution for classifying fingerspelled letters in images, such that after doing this we would attempt to create sentences.

Chapter 2

Related work

Absolventul va prezenta clar partea aplicativ? a lucr?rii ?i metodologia de solu?ionare folosind elementele teoretice.

Se va specifica mediul de lucru, a facilit??ilor folosite ?n acest mediu, proiectarea aplica?iei, detalii de implementare, exemple de test sau rezultate sub forma unor studii de caz, modul de utilizare a programului prin prezentarea documenta?iei de utilizare. Va fi anexat ?n lucrare inclusive codul surs?.

Partea dezvolt?rii aplicative poate fi constituit? din mai multe capitole.

Referirea unei figuri 2.1.

Figure 2.1: Ciclul de dezvoltare al sistemelor bazate pe componente adaptat modelului cascad?

Referirea la Tabelul 2.1.

Nume algoritm	Toate solu?iile	Solu?ia optim?
Nume 1	20	5
Nume 2	20	2

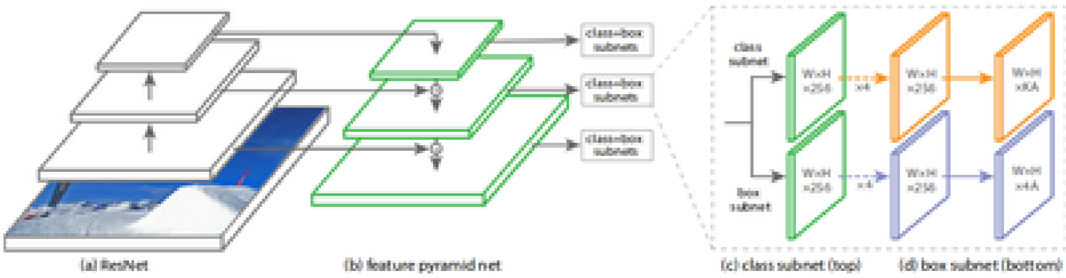
Table 2.1: Solu?ii ob?inute

Chapter 3

Proposed approach

3.1 Theoretical part

3.1.1 Architecture



The model our work is based on is RetinaNet. RetinaNet is a one stage detector having a convolutional neural network as a feature extractor for the 2 task-subnetworks it has. The first CNN's backbone is based on a Feature Pyramid Network build on a forward Residual Network. The FPN generates rich multi-scale feature maps. In [7] it is emphasised that we can also extract relevant features and make predictions based on feature maps that are on layers inside the network. The pyramid usually has 7 levels. The 2 task-subnetowrks are:

- The classification subnetwork is a FCN that appears on each level of the pyramid and tries to classify each anchor box and outputing a K-vetor of probabilities for each class(meaning we have K classes).
- The box regression subnetwork has the same architecture (but different parameters) like the classification subnetwork only outputing a 4-vector representing the targets of the bounding box.

3.1.2 The loss

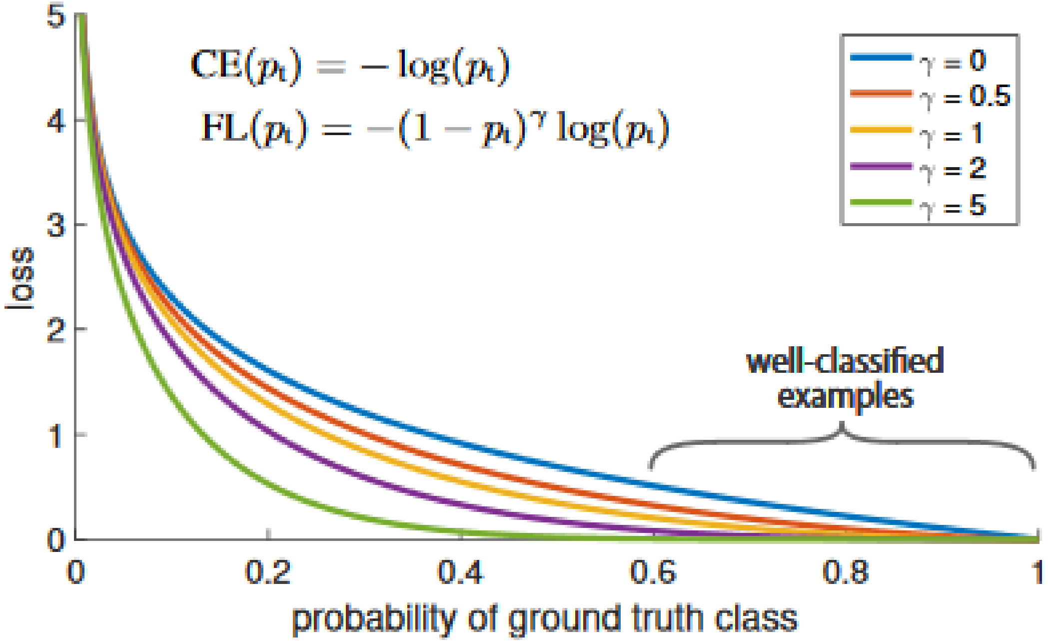


Figure 1. We propose a novel loss we term the *Focal Loss* that adds a factor $(1 - p_t)^\gamma$ to the standard cross entropy criterion. Setting $\gamma > 0$ reduces the relative loss for well-classified examples ($p_t > .5$), putting more focus on hard, misclassified examples. As our experiments will demonstrate, the proposed focal loss enables training highly accurate dense object detectors in the presence of vast numbers of easy background examples.

Instead of the cross-entropy loss function we used a more novel loss function called the focal loss. The reason behind using focal loss is that (instead of using Online Hard Example Mining) the loss emphasies learning from hard examples. The focal loss function stands at the end of the classification FCN and applied to all 100k anchor boxes sampled from a single image.

3.2 Application development

Chapter 4

Application

4.1 Methodology

I am Lord Voldemort Absolventul va prezenta detaliat (pe baza document?rii bibliografice problematica tratat?: soiangioionoinegoioinoin oinegwonai

- ?ncadrarea temei ?ntr-una mai general?;
- trecerea ?n revist? a abord?rilor existente ale problemei cu marcarea avantajelor ?i dezavantajelor;
- descompunerea ?n subprobleme specifice ?i prezentarea modului de rezolvare.

4.2 Dataset

4.3 Results

4.4 Discussion

SCRIU AICIIII Partea fundament?rii teoretice poate fi consituit? din mai multe capitole, ca de exemplu "Stadiul actual din domeniu/State of art/Literature Review" ?i "Modele teoretice ?i metode folosite/Research Method" ?i "Problem Statement".

Chapter 5

Conclusion and future work

Absolventul va realiza o autoevaluare a rezultatelor prezentate (punctarea aspectelor originale, a avantajelor ?i limitelor solutiilor oferite) ?i a eventualelor aspecte r?mase nerezolvate,

?n general se prezint? ?n urm?toarele subsec?iuni: Concluzii, Sumarul contribu?iilor, Direc?ii viitoare de cercetare.

5.1 Conclusion

5.2 Future work

Bibliography

[1] Ian Goodfellow, et al.. *Deep Learning.* MIT Press, Massachusetts, 2016.

[2] Lungociu Corneliu *Real Time Sign Language Recognition Using Artificial Neural Networks.* in Studia Univ. Babes-Bolyai, Informatica, Volume LVI, Number 4, 2011.

[3] Paranjape Ketki Vijay *Recent Developments in Sign Language Recognition: A Review.* in International Journal on Advanced Computer Engineering and Communication Technology (IJACECT), Volume-1, Issue-2, 2012.

[4] T.Starner and A. Pentland *Real-Time American Sign Language Recognition from Video Using Hidden Markov Models.* Computational Imaging and Vision, 9(1); 227-243, 1997.

[5] Mekala et al. *Real-Time Sign Language Recognition based on Neural Network Architecture.* System Theory (SSST), 2011 IEEE 43rd Southeastern Symposium 14-16 March 2011.

[6] Mekala et al. *Real-time American Sign Language Recognition with Convolutional Neural Networks.* Stanford Journal 2014.

[7] Ross Girshick and Ilija Radosavovic and Georgia Gkioxari and Piotr Dollar and Kaiming He. *Detectron* `https://github.com/facebookresearch/detectron`

[8] Mekala et al. *Real-time American Sign Language Recognition with Convolutional Neural Networks.* Stanford Journal 2014.