

# **Find objects in real estate images with convolutional neural networks**

OSCAR RÅHLÉN OCH SACHARIAS SJÖQVIST

Degree project in Computer Science

Date: March 23, 2019

Supervisor: Handledare

Examiner: Examinator

School of Electrical Engineering and Computer Science

Swedish title: Hitta object i fastighetsbilder med convolutional neural networks



## **Abstract**

English abstract goes here.

## **Sammanfattning**

Svenskt sammanfattning

# Contents

<b>1</b>	<b>Introduktion</b>	<b>1</b>
1.1	Problemformulering . . . . .	1
1.2	Frågeställning . . . . .	1
1.3	Avgränsningar . . . . .	1
<b>2</b>	<b>Bakgrund</b>	<b>2</b>
2.1	Maskininlärning . . . . .	2
2.1.1	Neural networks . . . . .	3
2.2	Convolutional Neural Network . . . . .	3
2.2.1	Arkitekturer . . . . .	3
2.3	Transfer Learning . . . . .	3
2.4	Fine-tuning och minska overfitting . . . . .	4
2.4.1	Data Augmentation . . . . .	4
2.4.2	Dropout . . . . .	4
<b>3</b>	<b>Metod</b>	<b>5</b>
<b>4</b>	<b>Resultat</b>	<b>6</b>
<b>5</b>	<b>Diskussion</b>	<b>7</b>
<b>6</b>	<b>Slutsats</b>	<b>8</b>
	<b>Bibliography</b>	<b>9</b>
<b>A</b>	<b>Appendix A</b>	<b>10</b>



# Chapter 1

## Introduktion

Information om ämnet, leda in läsaren samt förklara uppsatsens relevans [1].  
Vi borde sikta på ungefär 40 sidor totalt.

### 1.1 Problemformulering

Vad är syftet med uppsatsen? Vad vill ni uppnå

### 1.2 Frågeställning

Vår frågeställning

### 1.3 Avgränsningar

Vilka avgränsningar vi gjort i datamängder, testpersoner samt modeller.

# Chapter 2

## Bakgrund

### 2.1 Maskininlärning

En maskininlärningsalgoritm är en algoritm som kan lära sig och bli bättre från data [2]. Det kan beskrivas som ett datorprogram som kan följa en grupp med uppgifter  $T$  och lära sig med hjälp av erfarenhet  $E$  och dess prestanda  $P$  är mätbar. Maskininlärning hjälper oss att lösa problem som är för svåra för program skrivna och designade av människor att lösa.

Enligt [2] är ett vanligt problem att lösa med maskininlärning klassificeringsproblem. Det innebär att bestämma vilken av  $k$  klasser en viss indata tillhör. För att lösa detta skapar man vanligtvis en funktion som tar in indata  $x$  och returnerar  $y$ , vilket är ett tal som representerar klassen. Det går även att returnera en sannolikhetsdistribution över alla klasser. Enligt [2] så löses modern objektigenkänning med djupinlärning.

En typ av algoritmer inom maskininlärning är supervised learning algorithms [2]. En sådan algoritm lär sig av data (erfarenhet  $E$ ) som redan är uppmärkt med rätt klass. Detta kallas för labeled data. Det man vanligtvis gör är att man delar upp den märkta datan i ett träningsset och testset. Därefter tränar man algoritmen på sitt träningsset och testar sedan hur bra den fungerar, i form av fel (error) eller noggrannhet (accuracy), på sitt testset.

Enligt [2] så är så kallade AI-kompleta problem, som objektigenkänning, bra att ha en baseline modell som är baserad på djupinlärning. Det är även bra att börja med ett CNN om indata består av bilder.

Man borde använda någon form av regularization. Där är dropout bra och fungerar i många tillfällen. Early stopping borde nästan alltid användas. Batch normalization kan användas istället för dropout, för att minska generaliseringsfel.



Om problemet är studerat förut så kan det vara bra att utgå ifrån modeller och algoritmer som redan har visats sig fungera bra. I bildigenkänning så är det till och med vanligt att man utgår ifrån samma vikter, så kallad transfer learning. Vanligtvis tränas då CNN upp på bilder från ImageNet.

### 2.1.1 Neural networks

Neurala nätverk är en modell inom maskininläring. Dess mål är att approximera en funktion  $f$  [2]. De är en viktig del av maskininläringen och en specifik typ av feedforward network är convolutional neural networks, som är vanligt förekommande inom bildkategorisering.

Den nuvarande tekniken för objektigenkänning använder sig huvudsakligen av maskininläring [3]. För att öka dess prestanda så kan man samla in mer data, lära sig mer kraftfulla modeller eller använda sig av bättre tekniker för att undvika overfitting.

## 2.2 Convolutional Neural Network

Convolutional Neural Network är en typ av nätverk för indata som kan ses som ett rutnät [2]. I fallet om bilder så är det ett tvådimensionellt rutnät av pixlar. Det som gör CNN speciellt är att de använder sig av den linjära operationen convolution istället för vanlig matrismultiplikation i minst ett av sina lager.

För en så avancerad uppgift som objektigenkänning så räcker det inte med stora mängder data, det behövs en modell med tidigare erfarenhet (prior knowledge) som tar korrekta antaganden om bilder. Detta gör convolutional neural networks (CNNs) bra [3]. Jämfört med ett vanligt feedforward neural network med samma storlek på varje lager, så har CNNs färre anslutningar och färre parametrar, vilket gör att de blir enklare att träna.

### 2.2.1 Arkitekturer

Ett bra sätt att välja arkitektur är att testa flera olika arkitekturer och frysa alla lager förutom det sista, och se resultatet [2]. Efter man hittat den bästa så tränar man alla lager.

## 2.3 Transfer Learning

Transfer Learning handlar om att överföra information mellan två olika områden [4]. I datorseende (computer vision) handlar transfer learning huvudsak-

ligen om att minska träningsprocessen genom att använda ett neuralt nätverk som redan är tränat på andra kategorier av bilder. Så man överför vikterna från ett tränat nätverk till ett nytt nätverk och anpassar det sedan för att kategorisera det man anser att kategorisera.

## **2.4 Fine-tuning och minska overfitting**

Enligt [3] finns det minst två bra tekniker för att minska overfitting, data augmentation och dropout.

### **2.4.1 Data Augmentation**

### **2.4.2 Dropout**

# Chapter 3

## Metod

Hur vi gått tillväga. Vilka dataset, hur implementation gått till (verktyg, klassificerare, parametrar), hur vi valt features. Hur evalueringen har gått till (träning, test, validation set).

Vi vill mäta både precision och recall. Plotta en PR curve.

Vi kan även plotta hur precisionen har gått om i förhållande till mer data, för att skapa en uppfattning av om modellen kan bli bättre med mer data. Vanligtvis gör man detta i log-skala [2].

Gör även att göra en grid search på hyperparametrar.

Hur vi kan använda oss av preprocessing kan vi läsa i kapitel 12 av Goodfellow.

# Chapter 4

## Resultat

Presentation av resultatet från våra olika tekniker och evalueringar. i tabeller och grafer. Även beräkningstid.

# Chapter 5

## Diskussion

Diskutera resultatet och hur olika delar kan ha påverkat eller påverkade. Diskutera eventuell framtida forskning. Begränsningar med resultatet. Etiska aspekter. Hållbarhet.

# **Chapter 6**

## **Slutsats**

Slutsats av vad vi kom fram till.

# Bibliography

- [1] Werner Heisenberg and Paul Dirac. “To be or not to be”. In: *Proceedings of the Uncertain Society Annual Meeting*. Ed. by Erwin Schrödinger. 2015, pp. 21–22.
- [2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (visited on 03/22/2019).
- [4] Maxime Oquab et al. “Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks”. In: 2014, pp. 1717–1724. URL: [http://openaccess.thecvf.com/content\\_cvpr\\_2014/html/Oquab\\_Learning\\_and\\_Transferring\\_2014\\_CVPR\\_paper.html](http://openaccess.thecvf.com/content_cvpr_2014/html/Oquab_Learning_and_Transferring_2014_CVPR_paper.html) (visited on 03/22/2019).

**Appendix A**

**Appendix A**