

# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ  
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY  
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

## IMAGE CAPTIONING WITH RECURRENT NEURAL NETWORKS

SEMESTRÁLNÍ PROJEKT  
TERM PROJECT

AUTOR PRÁCE  
AUTHOR

Bc. JAKUB KVITA

BRNO 2015



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**  
BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA INFORMAČNÍCH TECHNOLOGIÍ**  
**ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ**

FACULTY OF INFORMATION TECHNOLOGY  
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

# **POPIS FOTOGRAFIÍ POMOCÍ REKURENTNÍCH NEU- RONOVÝCH SÍTÍ**

IMAGE CAPTIONING WITH RECURRENT NEURAL NETWORKS

**SEMESTRÁLNÍ PROJEKT**

TERM PROJECT

**AUTOR PRÁCE**

AUTHOR

**Bc. JAKUB KVITA**

**VEDOUcí PRÁCE**

SUPERVISOR

**Ing. MICHAL HRADIŠ, Ph.D.**

BRNO 2015

## **Abstrakt**

Výtah (abstrakt) práce v českém jazyce.

## **Abstract**

Výtah (abstrakt) práce v anglickém jazyce.

## **Klíčová slova**

Klíčová slova v českém jazyce.

## **Keywords**

Klíčová slova v anglickém jazyce.

## **Citace**

Jakub Kvita: Image Captioning with Recurrent Neural Networks, semestrální projekt, Brno, FIT VUT v Brně, 2015

# Image Captioning with Recurrent Neural Networks

## Prohlášení

Prohlašuji, že jsem tento semestrální projekt vypracoval samostatně pod vedením pana Michala Hradiše.

.....

Jakub Kvita  
December 20, 2015

## Poděkování

Zde je možné uvést poděkování vedoucímu práce a těm, kteří poskytli odbornou pomoc.

© Jakub Kvita, 2015.

*Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.*

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Neural networks</b>	<b>3</b>
2.1	Recurrent neural nets . . . . .	3
2.1.1	LSTM – Long Short-Term Memory . . . . .	3
2.1.2	GRU – Gated Recurrent Unit . . . . .	3
2.1.3	Processing text with RNN . . . . .	3
2.2	Convolutional neural nets . . . . .	4
<b>3</b>	<b>Experiments</b>	<b>5</b>
3.1	Frameworks . . . . .	5
3.1.1	Torch . . . . .	5
3.2	Predicting next character in sequence . . . . .	5
<b>4</b>	<b>Image captioning</b>	<b>6</b>
4.1	Related Work . . . . .	6
4.1.1	Show and Tell . . . . .	6
4.1.2	Show, Attend and Tell . . . . .	6
4.1.3	From Captions to Visual Concepts and Back . . . . .	6
4.2	Datasets . . . . .	6
4.2.1	MS COCO . . . . .	6
4.2.2	Flickr 30k,8k . . . . .	6
4.3	Evaluation metrics . . . . .	6
4.3.1	BLEU . . . . .	6
4.3.2	CIDEr . . . . .	6
4.3.3	METEOR . . . . .	6
<b>5</b>	<b>Model</b>	<b>7</b>
5.1	Architecture . . . . .	7
5.2	Training details . . . . .	7
<b>6</b>	<b>Conclusion</b>	<b>8</b>

# Chapter 1

## Introduction

Klasicky popis toho co se tady bude dit, jak je to dulezite, atd.

## Chapter 2

# Neural networks

General idea of neural networks was slowly emerging after World War II. Perceptron, as a single neuron unit, was created in 1958 by Frank Rosenblatt<sup>1</sup>, but became popular only after creation of backpropagation algorithm in 1975. At that time neural nets have not reached massive popularity, not because they are not working, but due to small computing power of machines back then and lack of datasets. Recently (after 2000) neural nets became popular again. Mostly because researchers dealt with the problems from before and successfully applied neural nets in multiple fields like computer vision, speech recognition and natural language processing.

Since then various useful architectures and algorithms are now introduced almost every month. There is vast amount of various architectures and algorithms, in this chapter, I will describe only a couple – those, which are used in this thesis.

### 2.1 Recurrent neural nets

Nekde tady nacpat rozbor clanku o rekurentnich sitich?

Zduraznit problem vanishing a exploding gradientu

Popis toho jak umi pracovat se sekvencema, predikci dalsiho prvku, da se pouzit na spoustu veci, zvuky, ceny na burze, preklady, predikci textu.

#### 2.1.1 LSTM – Long Short-Term Memory

Jak to vyresilo problem vyse. Pridat i rovnice, ktere pouzivam ja, rozebrat dopodrobna.

#### 2.1.2 GRU – Gated Recurrent Unit

Zminit jako updatovanou verzi

#### 2.1.3 Processing text with RNN

Mozna trochu upravit nazev. (Character level and word level embeddings)

<sup>1</sup>The perceptron: A probabilistic model for information storage and organization in the brain. Rosenblatt, F. Psychological Review, Vol 65(6), Nov 1958, 386–408.

Popis toho jak se pracuje s textem v rnn, ze to je taky sekvence. Character level, word level, embeddings. Popis rozdilu toho jak funguji preklady a generovani dalsiho prvku sekvence.

## 2.2 Convolutional neural nets

Kratky uvod do toho, kde se používají, jak se vyvinuly, jednoduchy popis toho jak funguji. Obrazek?

Asi neni potreba davat subsekcce na vrstvy, staci popsát jak to funguje vsechno dohromady, jednotlivé vrstvy ve větách v jednom odstavci. Obrazek.



## Chapter 3

# Experiments

Kapitola jen na semestralni projekt. V diplomce ji odstranim.

### 3.1 Frameworks

Jak se to implementuje, jake knihovny se pouzivaji.

Cast o Torchi se zrecykluje do diplomky.

Popsat ze popularni je Caffe, ukladani modelu v caffe, theano, pythoni veci zalozene na theano, Torch, TensorFlow.

#### 3.1.1 Torch

Udelat tady tabulku o ruznych balicich co torch ma

### 3.2 Predicting next character in sequence

Jak jsem to udelal, co to dela, ukazky.

Reference na Karpathyho char-rnn

# Chapter 4

## Image captioning

Znovu uvod k tomu jak je to dulezite a tentokrat jak na tom lidi pracuji, co je potreba a jak se to hodnoti.

### 4.1 Related Work

Dat tomu nejake lepsi jmeno, clanky o popisovani obrazku ktere jsem cetl, pouzil.

#### 4.1.1 Show and Tell

#### 4.1.2 Show, Attend and Tell

#### 4.1.3 From Captions to Visual Concepts and Back

### 4.2 Datasets

COCO, Flickr, popis jake jsou.

#### 4.2.1 MS COCO

#### 4.2.2 Flickr 30k,8k

### 4.3 Evaluation metrics

BLEU, cIDER, jak se používají, co dělají...

#### 4.3.1 BLEU

#### 4.3.2 CIDEr

#### 4.3.3 METEOR

# Chapter 5

## Model

Do semestrálního projektu nebo až na diplomku?

Design modelu, co chci použít, jaké metody chci zkusit.

### 5.1 Architecture

Architektura modelu, jaké matematické modely jsem použil, bez implementačních detailů.

### 5.2 Training details

Popis pomocí jakého algoritmu jsme trénovali, s jakými parametry, minibatches, datasety.

## Chapter 6

# Conclusion

Udelat jeden zaver pro semestralni projekt, pak ho prepsat pro diplomku.

# Bibliography

- [1] Kolektiv autorů. *Pravidla českého pravopisu*. Academia, 2005. ISBN 80-200-1327-X.
- [2] Jiří Rybička. *L<sup>A</sup>T<sub>E</sub>X pro začátečníky*. Konvoj, 1999. ISBN 80-85615-77-0.
- [3] Z. Rábová, P. Hanáček, P. Peringer, P. Přikryl, and B. Křena. Užitečné rady pro psaní odborného textu [online].  
[http://www.fit.vutbr.cz/info/statnice/psani\\_textu.html](http://www.fit.vutbr.cz/info/statnice/psani_textu.html), 2008-11-01 [cit. 2008-11-28].