Performance Evaluation of GANs in a semi-supervised OCR Use Case
Dr. Florian Wilhelm
Principal Data Scientist @ inovex

Special Interests
• Mathematical Modelling
• Recommendation Systems
• Data Science in Production
• Python Data Stack
• Maintainer of PyScaffold

@FlorianWilhelm
FlorianWilhelm
florianwilhelm.info
IT-project house for digital transformation:

- Agile Development & Management
- Web · UI/UX · Replatforming · Microservices
- Mobile · Apps · Smart Devices · Robotics
- Big Data & Business Intelligence Platforms
- Data Science · Data Products · Search · Deep Learning
- Data Center Automation · DevOps · Cloud · Hosting
- Trainings & Coachings

Using technology to inspire our clients. And ourselves.
Agenda

1. Use Case
2. Text Spotting
3. Data and Pipeline
4. Generative Adversarial Networks
5. Semi-supervised Learning
6. Results
Vehicle Identification Number (VIN)

Unique identifier like a fingerprint of a vehicle

Use Case

Spotting the vehicle identification number (VIN) in images of vehicle registration documents

Information about the car:

- Manufacturer: BMW
- Model: X3
- Year: 2013-03-21
- Engine power: 143 PS
- Equipment: - Xenon Lights ...

VIN: WF0DXXGAKDEJ37385
OCR - Libraries

<table>
<thead>
<tr>
<th>Commercial software</th>
<th>Open source tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Vision API</td>
<td>Tesseract</td>
</tr>
<tr>
<td>ABBYY</td>
<td>PyOCR</td>
</tr>
<tr>
<td>Adobe</td>
<td>GOOCR</td>
</tr>
<tr>
<td>nitro</td>
<td></td>
</tr>
</tbody>
</table>
OCR with Tesseract
Agenda

1. Use Case
2. Text Spotting
3. Data and Pipeline
4. Generative Adversarial Networks
5. Semi-supervised Learning
6. Results
Methodology in Text Spotting

Girshick et al. (2014), “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation”

Character detection & extraction

- Connected components
- Stroke width transform
- Edge detection

- SVM
- Learning with HOG
- CNN

- Region proposal
- Hypotheses CNN pooling

Character recognition

- SVM
- Nearest Neighbor
- CNN
- CNN + RNN
- ...
Convolutional Neural Network

Convolution with 3x3 kernel and stride = 1

Max pooling with a 2x2 filter and stride = 2
Agenda

1. Use Case
2. Data and Pipeline
3. Semi-supervised Learning
4. Generative Adversarial Networks
5. Semi-supervised Learning
6. Results
Objectives

Dataset:

- ~170 images of vehicle registration documents

1. Implementation of a prototype

2. Comparison of classifiers

Text Spotting

„XLG0H200NA0A10348“

a) Supervised method

b) Semi-supervised method
End-to-End Text Spotting Pipeline

Region of Interest Extractor

Sliding window

Character Detector (2 classes)

Non Maximum Suppression

Character Recognizer (36 classes)

Image depicting only VIN

All windows

All windows with characters

Only one window per character
Small Dataset
What to do about that?

1. Data Generation

2. Data Augmentation
Data Augmentation

Datasets:

2 classes

36 classes

Data augmentation:

Original image labeled manually as "0"

Character Detector (2 classes)

Character Recognizer (36 classes)

Label: "character"

Label: "no character"

Label: "0"
Datasets

170 images of vehicle registration documents

Training set

85 images

Testing set

85 images

Data Augmentation

Training sets of classifiers

Detector

~ 42000 images
2 classes

Recognizer

~ 8000 images
36 classes

Testing sets of classifiers

Detector

~ 42000 images
2 classes

Recognizer

~ 8000 images
36 classes

Testing sets of pipeline

~ 8000 images
36 classes

85 images
Classifiers

1. Supervised Convolutional Neural Network

2. Semi-supervised Generative Adversarial Network
Agenda

1. Use Case
2. Text Spotting
3. Data and Pipeline
4. Generative Adversarial Networks
5. Semi-supervised Learning
6. Results
Yann LeCun
Director of Facebook AI Research, Prof at NYU

“... (GANs) and the variations that are now being proposed is the most interesting idea in the last 10 years in ML, in my opinion.”

Ian J. Goodfellow @ Google Brain
Generative Adversarial Network

Goal: Generate images, which seem to be realistic

Goal: Differentiate between fake and real images
Generative Adversarial Network

Real labeled images

"D classified the generated image as 10% real"

"Yes"

Is D correct?
Mathematical formulation

Objective function

Discriminator calculates likelihood [0,1] for an image being real

\[
\min_{\theta_g} \max_{\theta_d} \left[ \mathbb{E}_{x \sim p_{\text{data}}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \right]
\]

Discriminator output for real images

Discriminator output for fake images

Training (alternating)

Maximizing discriminator loss

\[
\max_{\theta_d} \left[ \mathbb{E}_{x \sim p_{\text{data}}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \right]
\]

Minimizing generator loss

\[
\min_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z)))
\]
Example of generated images

Training images:

Generated images during learning process:
Agenda

1. Use Case
2. Text Spotting
3. Data and Pipeline
4. Generative Adversarial Networks
5. Semi-supervised Learning
6. Results
Semi-supervised Learning

- Makes use of unlabeled data
- Combines supervised and unsupervised learning
Semi-supervised GAN for Character Detection

Real labeled images

Real unlabeled images

Generator

Discriminator
Agenda

1. Use Case
2. Text Spotting
3. Data and Pipeline
4. Generative Adversarial Networks
5. Semi-supervised Learning
6. Results
Character Detector (2 classes)

Pretraining of DCNN
Manually generated images with CAPTCHA methods

- "Character"
- "No character"

Accuracy vs. Size of labeled training set
Character Detector (2 classes)

![Graph showing accuracy vs size of labeled training set for different models: DCNN, DCNN pretrained, and Supervised GAN. The graph indicates that Supervised GAN achieves the highest accuracy for a given size of labeled training set.]
Character Detector (2 classes)

![Character Detection Graph](image)

**Accuracy vs. Size of Labeled Training Set**

- DCNN
- DCNN pretrained
- Supervised GAN
- Semi-supervised GAN

**Semi-supervised GAN**

- Real labeled images
- Real unlabeled images

**Diagram**

- Discriminator
- Generator

**Legend**

- C
- F
Character Recognizer (36 classes)

Accuracy vs Size of labeled training set for different models:
- DCNN
- DCNN pretrained
- Supervised GAN
- Semi-supervised GAN
End-to-End Text Spotting Pipeline

Region of Interest Extractor

Sliding window

Character Detector (2 classes)

Non Maximum Suppression

Character Recognizer (36 classes)

85 images

1. WWWZZZAUZFW337684
2. WMXNM310GCTC69478

... 85. VFLKWX82550717152

Accuracy = 99.94%
Google Cloud Vision API vs. Our Approach

Google Cloud Vision API

85 images of VINs

Google Cloud Vision API

∅ Levenshtein distance = 4.49

Character Detector (2 classes)

Non Maximum Suppression

Region of Interest Extractor

Character Recognizer (36 classes)

85 images

Levenshtein distance:

Classification

Label

AYZ33

XYZ321

= 3

∅ Levenshtein distance = 0.011
Key Learnings

- Custom solutions can tremendously outperform off-the-shelf software in a specific use-case
- Semi-supervised GANs can be successfully applied in use-cases with little data
- With simple data augmentation techniques having only little data might be enough
Bibliography

- Krizhevsky et al. (2012) „ImageNet Classification with Deep Convolutional Neural Networks“
- Girshick et al. (2014), „Region-Based Convolutional Networks for Accurate Object Detection and Segmentation“
- Girshick et al. (2015), „Fast R-CNN“
- Girshick et al. (2015), „Faster R-CNN“
- He et al. (2017), „Mask-R-CNN“
- Goodfellow et al. (2014) „Generative Adversarial Networks“
Thank you!

Florian Wilhelm
Principal Data Scientist

inovex GmbH
Schanzenstraße 6-20
Kupferhütte 1.13
51063 Köln
florian.wilhelm@inovex.de