



HOCHSCHULE LANDSHUT

HOCHSCHULE FÜR ANGEWANDTE WISSENSCHAFTEN

FAKULTÄT INFORMATIK

Expose der Masterarbeit

GENERIERUNG ZUSÄTZLICHER TRAININGSDATEN MITTELS
UNPAIRED IMAGE-TO-IMAGE TRANSLATION ZUR
VERBESSERUNG DER ERKENNUNGSLEISTUNG EINES MODELLS
ZUM ABLESEN VON STROMZÄHLERWERTEN

Andreas Filinger

Problemstellung

Ziel der Arbeit ist es die Erkennungsleistung eines Neuronalen Netzes zum Ablesen von Stromzählerwerten zu verbessern, indem neue Trainingsdaten durch unpaired Image-to-Image Translation generiert werden.

Die Leistung eines Neuronalen Netzes hängt ab von der verwendeten Trainingsdatenmenge. Diese sollte möglichst groß sein und das zu lösende Problem gut abdecken. Doch nicht immer ist eine solche Datenmenge gegeben: Die Beschaffung von Daten bestimmter Probleme kann zu aufwendig für eine große Datenmenge sein, oder die Verteilung der Klassen innerhalb der Datenmenge kann ungleichmäßig sein. Der Begriff „Data Augmentation“ fasst Methoden zusammen die den Trainingsdatensatz erweitern, indem vorhandene Daten verarbeitet werden. Um die Robustheit des Modells zu erhöhen können beispielsweise Bilder gedreht, gespiegelt oder mit Rauschen versehen werden. Zur Ausbalancierung der Klassenverteilung können Dateninstanzen einer unterrepräsentierten Klasse dupliziert werden. Eine weitere Möglichkeit Trainingsdaten zu erweitern bieten Generative Networks. 2017 beschäftigten sich Wang et al [WMYS17] mit der Erkennung von KFZ-Nummernschildern und lösten das Problem ungleichmäßiger Klassenverteilung in ihren Trainingsdaten (d.h. bestimmte Zeichen kommen in realen Nummernschildern häufiger vor als andere). Dazu verwendeten sie CycleGAN [ZPIE17], eine Variation von Generative Adversarial Networks, um synthetische Nummernschilder mit selbstgewähltem Inhalt in realistische Nummernschilder zu überführen.

Die Ergebnisse dieses Papers legen nahe, solch eine Herangehensweise auch auf andere Probleme anzuwenden. Ein Anwendungsfall mit ähnlichen Problemen ist das Ablesen von Stromzählerwerten mithilfe einer Kamera: Auch hier weisen Trainingsdaten (Bilder realer Stromzähler) eine ungleichmäßige Klassenverteilung auf (bestimmte Ziffern sind häufiger als andere). Es stellt sich die Frage, ob die Generierung zusätzlicher Trainingsdaten durch unpaired Image-to-Image Translation Techniken wie CycleGAN auch für dieses Problem eine Steigerung der Erkennungsleistung bewirken kann.

Zielsetzung

Als Ergebnis der Arbeit soll eine Pipeline entstehen, die als Input eine beliebige Ziffernfolge nimmt und daraus ein realistisches Bild eines Stromzählers generiert. Das Ziel ist es mithilfe von unpaired Image-to-Image Translation Bilder zu generieren, die, wenn als zusätzliche Trainingsdaten verwendet, die Erkennungsleistung eines Modells zum Ablesen von Stromzählerwerten steigern. Die Erfüllung dieses Ziels kann anhand eines Testdatensatzes wie UFPR-AMR [LBD⁺19], einer Sammlung von Stromzählerbildern, gemessen werden. Hierzu vergleicht man die Erkennungsleistung zweier sonst identischer Modelle, von denen eines zusätzlich mit den generierten Daten trainiert wurde.

Der erhoffte Nutzen der Arbeit, sollte dieses Ziel erreicht werden können, ist es die Realisierung eines Gerätes zum Ablesen von Stromzählerwerten zu unterstützen. Das Gesetz verlangt, dass bis 2032 in jedem Haushalt ein Stromzähler verbaut ist, der in der Lage ist automatisch Zählerwerte abzulesen und über das Internet zu versenden. [Bun] Die Kosten für die Anschaffung eines solchen Smartmeters würde der Verbraucher selbst tragen. Eine Alternative zum Austauschen alter Zähler könnte ein Gerät sein, welches mit einer fest montierten Kamera ein Bild des Zählers macht, daraus den Zählerstand mithilfe eines Neuronalen Netzes auf einem Prozessor ermittelt und schließlich über eine Kommunikationseinheit versendet.

Konkret ergeben sich zur Realisierung der Pipeline folgende Aufgaben:

- Recherche und Auswahl aktueller unpaired Image-to-Image Methoden
- Implementierung eines Bildgenerators, der aus Ziffernfolgen synthetische Zählerstandbilder erzeugt
- Implementierung einer (oder mehrerer) unpaired Image-to-Image Methode, die synthetische Zählerstandbilder in realistische Bilder überführt
- Evaluierung mithilfe eines Benchmarks

Stand der Forschung

Mit GANs [GPAM⁺14] können beginnend mit einem Anfangsdatensatz weitere (realistische) Bilder erzeugt werden. Dies wird durch die Verwendung von zwei Neuronalen Netzen erreicht, die jeweils gegensätzliche Ziele verfolgen: Der Generator versucht möglichst realistische Bilder zu erzeugen, während der Diskriminator versucht reale Bilder von generierten zu unterscheiden. Auf diese Art erzeugte Bilder wurden bereits dazu verwendet, zusätzliche Trainingsdaten für Neuronale Netze zu erzeugen. Yi et al [YWB19] sammelten in Ihrer Studie Anwendungen von GANs im medizinischen Bereich. GANs sind hier besonders attraktiv, da Trainingsdaten oft schwer zu beschaffen sind, etwa weil die Privatsphäre von Patienten betroffen ist. So steigerten beispielsweise Frid-Adar et al [FADK⁺18] die Erkennungsleistung verschiedener Leberläsionen, indem sie zu jeder der drei Klassen synthetische Bilder mit GANs erzeugten.

CycleGAN [ZPIE17] ermöglicht eine Abbildung von einer Bildmenge (Domäne) in eine andere, etwa ein Foto einer Landschaft im Sommer in ein Foto im Winter. Image-to-Image Translation gab es zwar schon vor CycleGAN, bisher benötigte man aber eine Menge korrespondierende Bildpaare beider Domänen, zum Beispiel jeweils die selbe Landschaft im Sommer und im Winter (paired Image-to-Image Translation). Bei CycleGAN benötigt man keine korrespondierenden Bildpaare mehr, es reichen zwei Mengen unabhängiger Bilder beider Domänen (unpaired Image-to-Image Translation).

Wie oben erwähnt verwendeten Wang et al [WMYS17] CycleGAN, um synthetische Bilder von Nummernschildern in realistische zu überführen. Damit generierten sie eine Menge realistischer Trainingsdaten, deren Inhalt (Ziffernfolge) vorgegeben werden konnte. Diese Herangehensweise wird im Rahmen dieser Masterarbeit auf das Problem der Erkennung von Stromzählerwerten angewendet.

Vorarbeiten/Vorkenntnisse

Eine Einführung in das Thema Machine Learning bietet die Vorlesung „Vertiefung Datenbanksysteme“ im zweiten Mastersemester. Dort wurden Grundlegende Begriffe und Verfahren behandelt, es wurden beispielsweise Algorithmen wie die Lineare Regression, das 1-Regel-Verfahren und der Prism-Algorithmus vorgestellt. Diese wurden im Praktikum mithilfe des Machine-Learning Tools WEKA [Uni] angewendet.

Als Einstieg in Neuronale Netze behandelte die Coursera-Spezialisierung „Deep Learning“ [Ng], eine Menge von Onlinekursen welche mittels Videos Wissen vermittelt, eine Vielfalt von Themen: Beispiele reichen von Aufbau, Funktionsweise und dahinterliegenden Algorithmen Neuroner Netze, über Optimierungen von Gradient Descent und Strategien zum Aufbau eines Deep Learning Projekts (Aufteilung der Daten in Train/Dev/Test; Beschreibung veränderbarer Hyperparameter) bis hin zu Convolutional Networks. Auch wurden einige Beispiele Neuroner Netze vorgestellt, etwa ResNet oder Inception Network.

Einen Überblick über Generative Modelle bieten je eine (auf Youtube verfügbare) Vorlesung zu diesem Thema vom MIT [Ami] und Stanford [Sta], sowie zwei vom Erfinder von GANs gehaltene Präsentationen [Goob] [Gooa] (ebenfalls auf Youtube verfügbar).

Zum Praktischen Einstieg erfolgte die Einarbeitung in das Machine-Learning Framework Keras [Ker]. Angefangen mit einer typischen „Hello World“-Anwendung, welche handgeschriebene Ziffern des MNIST-Datensatzes erkennt, wurde insbesondere ein Beispiel einer GAN-Anwendung [gan] implementiert, welche MNIST-Ziffern generiert.

Methoden

Zur Erfüllung der obengenannten Ziele ergeben sich folgende Aufgaben:

Recherche aktueller unpaired Image-To-Image Translation Methoden

CycleGANs wurden erstmals 2017 vorgestellt, seitdem sind zahlreiche weitere unpaired Image-To-Image Translation Methoden erschienen. Es muss eine Suche und Auswahl einer (bzw. mehrerer) dieser Methoden durchgeführt werden, die im Rahmen dieser Arbeit implementiert werden. Hierzu können moderne Suchmaschinen wie semanticscholar oder google scholar verwendet werden.

Datenbeschaffung

Sowohl für die Image-To-Image Translation als auch für die Klassifizierung werden Trainingsdaten, Bilder von Stromzählern, benötigt. Einerseits bietet der UFPR-AMR Datensatz 2000 Bilder. Sollte sich diese Datenmenge als zu gering herausstellen, wäre auch eine Umfrage denkbar, bei der Studenten der Hochschule Landshut per Rundmail um ein mit dem Smartphone geschossenes Foto ihres Stromzählers gebeten werden.

Erstellung eines Generators synthetischer Zählerstandbilder

Der erste Teil der obenbeschriebenen Pipeline ist ein Bildgenerator, der aus einer beliebigen Ziffernfolge ein synthetisches Bild eines Stromzählers generiert. Diese Bilder haben nicht den Anspruch realistisch sein, und können also durch Skripts generiert werden, etwa mit openCV oder als SVG-Bild.

Implementierung einer unpaired Image-To-Image Translation Methode

Die gewählte Image-To-Image Translation Methode soll implementiert und jeweils mit generierten synthetischen und realen Bildern aus den Trainingsdaten trainiert werden, sodass das Modell synthetische Bilder in realistische überführt.

Evaluation

Mit der fertigen Pipeline können aus beliebigen Ziffernfolgen realistische Bilder von Stromzählern generiert und damit der bestehende Datensatz erweitert werden. Die Qualität hiervon lässt sich dann bestimmen, indem man zwei Instanzen desselben Klassifizierers vergleicht: Eines trainiert man nur mit den ursprünglich vorhandenen Daten, das andere zusätzlich mit den generierten. Idealerweise weist das Modell, welches zusätzlich mit generierten Daten trainiert wurde, eine höhere Erkennungsrate auf. Im ersten Anlauf wird aber eher erwartet, dass das Ergebnis nicht so positiv ausfällt. Deshalb seien obige Aufgaben nicht als einmalige Arbeitsschritte anzusehen, sie sollen eher in einem iterativem Prozess abgearbeitet werden. Ist das Endergebnis nicht zufriedenstellend, sollen sowohl die Auswahl der verwendeten Image-To-Image Translation Methode, als auch die Trainingsdaten, der Generator synthetischer Bilder und die Implementierung der Image-To-Image Methode überdacht werden.

Metriken

Um die Qualität generierter Bilder während der Entwicklung zu messen wird die Metrik FID (Fréchet Inception Distance) [HRU⁺17] eingesetzt: FID ist eine Metrik, welche die Qualität generativer Modelle misst, indem die Distanz der Menge generierter Bilder und der Menge realer Bilder berechnet wird. Hierzu wird ein vortrainiertes Bildklassifizierungsmodell, das Inception Network, verwendet. Je geringer der Wert von FID, desto ähnlicher sind die generierten Bilder den realen Bildern.

Weiterhin kann die Diversität generativer Modelle durch die Metrik LPIPS (Learned Perceptual Image Patch Similarity) [ZIE⁺18] gemessen werden. LPIPS bietet eine Distanzmetrik zwischen einzelnen Bildern. Um die Diversität eines generativen

Modells zu messen berechnet man zuerst zu einem Input des generativen Modells mehrere Outputs. Diese Outputs werden dann durch LPIPS verglichen und der Durchschnitt hieraus gebildet. Dies wiederholt man dann für mehrere Inputs und bildet einen Durchschnitt, welcher dann die Diversität des Modells wiedergibt.

Zeitplan

Aus obigen Aufgaben ergeben sich folgende Meilensteine:

1. Die Recherche aktueller unpaired Image-to-Image Translation Methoden ist abgeschlossen, eine oder mehrere Methoden wurden ausgewählt. (2 Wochen)
2. Die Implementierung des Generators synthetischer Zählerstandbilder ist abgeschlossen. (1 Monat)
3. Die gewählte Image-to-Image Translation Methode wurde implementiert, damit ist die erste Version der Pipeline fertig. (1 Monat)
4. Nach wiederholter Iteration: Entweder ist das Ergebnis zufriedenstellend (d.h. die Generierung neuer Trainingsdaten führt zu gesteigerter Erkennungsleistung), oder es konnte trotz wiederholter Iteration keine Steigerung der Erkennungsleistung erzielt werden. (3 Monate)
5. Die Arbeit wird abgegeben.

Literaturverzeichnis

- [Ami] Alexander Amini. MIT 6.S191 (2019): Deep Generative Modeling. youtube.com/watch?v=yFBF11cLYx8. Letzter Zugriff: 2020-03-22.
- [Bun] Bundesnetzagentur. Moderne messeinrichtungen/intelligente messsysteme. https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Verbraucher/Metering/SmartMeter_node.html. Letzter Zugriff: 2020-03-23.
- [FADK⁺18] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018.
- [gan] Tutorial: Deep convolutional generative adversarial network. tensorflow.org/tutorials/generative/dcgan. Letzter Zugriff: 2020-03-22.
- [Gooa] Ian Goodfellow. Aaai-19 invited talk - ian goodfellow (google ai) - adversarial machine learning. youtube.com/watch?v=AJJRWFVfNPg. Letzter Zugriff: 2020-03-22, hochgeladen von: Thanh Nguyen Nguyen.
- [Goob] Ian Goodfellow. Ian goodfellow: Generative adversarial networks (nips 2016 tutorial). <https://www.youtube.com/watch?v=HGYEUSm-0Q>. Letzter Zugriff: 2020-03-22, hochgeladen von: Steven Van Vaerenbergh.
- [GPAM⁺14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial networks. *ArXiv*, abs/1406.2661, 2014.
- [HRU⁺17] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. 12 2017.

- [Ker] Keras: The Python Deep Learning library. <https://keras.io/>. Letzter Zugriff: 2020-03-22.
- [LBD⁺19] R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz, and D. Menotti. Convolutional neural networks for automatic meter reading. *Journal of Electronic Imaging*, 28:1–14, 2019.
- [Ng] Andrew Ng. Coursera Spezialisierung Deep Learning. <https://www.coursera.org/specializations/deep-learning>. Letzter Zugriff: 2020-03-22.
- [Sta] Stanford University School of Engineering. Lecture 13 — generative models. youtube.com/watch?v=yFBF11cLYx8. Letzter Zugriff: 2020-03-22.
- [Uni] University of Waikato. WEKA - the workbench for machine learning. <https://www.cs.waikato.ac.nz/ml/weka>. Letzter Zugriff: 2020-03-22.
- [WMYS17] Xinlong Wang, Zhipeng Man, Mingyu You, and Chunhua Shen. Adversarial generation of training examples: Applications to moving vehicle license plate recognition. 2017.
- [YWB19] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552, 2019.
- [ZIE⁺18] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.
- [ZPIE17] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.