

Abgabe Digitale Bildverarbeitung

Simon Ball

Matrikelnummer: 3326441

Kurs: TFE21-2

5. Januar 2024

Die genutzten Daten stammen aus der MNIST-Datenbank für handgeschriebene Ziffern.¹ Alle in folgenden Abschnitten gezeigten Konzepte sollen, soweit möglich, die Anzahl der Labels und die Modellkomplexität (Modellparameter) minimieren sowie die Genauigkeit maximieren. Um Vergleichbarkeit zu schaffen, wird der reduzierte Datensatz von 1000 gelabelten Bildern für alle Konzepte verwendet. Mit dem Werkzeug **Weights and Biases** wurden jeweils verschiedene Hyperparameterkombinationen und deren Auswirkung auf den Trainingsprozess verglichen. Zusätzlich wird bei jedem Training die Lernrate dynamisch angepasst, sowie das beste Modell des gesamten Trainings durch entsprechende Callbacks gespeichert.

1 Konzepte

1.1 Data Augmentation

Um die Anzahl der gelabelten Daten reduzieren zu können und dabei trotzdem eine hohe Genauigkeit zu behalten, wurde Data Augmentation mit der Bibliothek **Augmentor** genutzt. Diese ermöglicht das verändern der Bilder durch Rotation, Zoom oder Verzerrung. Die entsprechenden Parameter wurden experimentell ermittelt und durch optische Sichtung einiger Beispiele überprüft, sodass die Ziffern noch gut erkennbar bleiben. Außerdem wird dadurch Overfitting verhindert.

¹Yann LeCun, Corinna Cortes und CJ Burges. “MNIST handwritten digit database”. In: *ATT Labs [Online]*. Available: <http://yann.lecun.com/exdb/mnist> 2 (2010).

1.2 Convolutional Neural Network

Das erste angewendete Modellart ist ein Convolutional Neural Network. Als Basis wurde ein online veröffentlichtes Modell² verwendet und weiter angepasst. Darunter fällt das Herauslöschten einer Dense- sowie Dropout-Schicht, wodurch sich die Anzahl der Parameter fast geviertelt haben, ohne die Genauigkeit signifikant zu verschlechtern. Durch ein Hinzufügen einer BatchNormalization- und ReLU-Aktivierungsschicht wurde die Genauigkeit ohne Hinzufügen von Parametern um etwa einen Prozentpunkt angehoben. Somit liegt die schlussendliche Genauigkeit bei etwa 94 % mit etwa 49.000 Parametern.

1.3 Monte Carlo Dropout

Bei Monte Carlo Dropout geht es um die Schätzung der Unsicherheit von Modellen. Durch das Anwenden von Dropout während der Testphasen und mehrmaliges Vorhersagen von Ergebnissen, kann das Modell eine Verteilung der Vorhersagen ausgeben.³

Das wird stark vereinfacht im Notebook umgesetzt, indem die Ergebnisnummer mit der höchsten Wahrscheinlichkeit nach mehreren Durchläufen als Ergebnis gewertet wird. Diese Technik wurde bei dem Convolutional Neural Network angewendet, wodurch die Genauigkeit um einen Prozentpunkt erhöht werden konnte.

1.4 Neural Network mit vorgeschalteter Hauptkomponentenanalyse

Eine weitere Möglichkeit die Bilder zu verarbeiten ist es, zunächst die Anzahl der Dimensionen zu reduzieren. Da jedes Bild 28x28 Pixel hat, kann es als 784-dimensionaler Eingangswert für ein Neuronales Netz angesehen werden. Um die Anzahl der Dimensionen zu reduzieren kann die Hauptkomponentenanalyse (PCA) genutzt werden. Dafür wurde experimentell bestimmt, dass die optimale Ergebnisdimension bei etwa 50 liegt. Da dieser Ansatz keine Labels benötigt, werden für die PCA die gesamten verfügbaren Trainingsdaten genutzt. Nur für das Training werden die Daten wieder auf 1000 reduziert. Als neuronales Netz wird anschließend eine Kombination aus zwei Dense-Schichten genutzt. Dieser Wert wurde experimentell ermittelt. Dennoch konnte nur eine Genauigkeit von ca. 83 % mit 3914 Parametern (PCA ausgeschlossen) erreicht werden.

²<https://www.kaggle.com/code/lakhindr/small-efficient-neural-network-for-mnist>

³Yarin Gal und Zoubin Ghahramani. *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*. 2016. arXiv: 1506.02142 [stat.ML].

1.5 Transfer Learning

Mit Transfer Learning besteht die Möglichkeit, Wissen aus einem trainierten Modell für eine andere Aufgabe wiederverwenden kann. In diesem Fall wurde das Modell ResNet50 mit den vortrainierten Gewichten des Datensatzes ImageNet verwendet, die während des Trainings nicht mehr verändert werden. Nur zwei zusätzlich hinzugefügte Schichten werden trainiert. Trotz der großen Anzahl von über 23 Millionen Parametern konnte nur eine Genauigkeit von etwa 80 % erreicht werden.

2 Fazit

Am schlechtesten abgeschnitten hat der letzte Transfer Learning Ansatz mit einer vergleichsweise hohen Modellkomplexität und dennoch geringen Genauigkeit. Das könnte daran liegen, dass das Netz für vorgegebene Aufgabe zu komplex ist und somit kein gutes Ergebnis erzielen kann. Bei der vorgeschalteten PCA konnte dagegen mit einem Bruchteil der Parameter ein ähnliches Ergebnis erzielt werden. Falls ein sehr kleines Netz gewünscht ist und die gegebene Genauigkeit ausreicht, kann dieses Modell gewählt werden. Im Bezug auf die Genauigkeit hat das Convolutional Netz mit Monte Carlo Dropout am besten abgeschnitten. Da letzteres durch eine große Anzahl an Vorhersagen errechnet wird, muss für die jeweilige Anwendung entschieden werden, ob die gewonnene Genauigkeit die zusätzliche Rechenzeit rechtfertigt.

Durch ein Erhöhen der Anzahl an Labels kann bei allen drei Netzen die Genauigkeit weiter verbessert werden.