Artificial intelligence and the automatic classification of historical photographs

Florian Eiler
Technische Hochschule Deggendorf
Grafenauer Straße 22, 94078 Freyung
Germany

Simon Graf
Technische Hochschule Deggendorf
Grafenauer Straße 22, 94078 Freyung
Germany
simon.graf@th-deg.de

Wolfgang Dorner
Technische Hochschule Deggendorf
Grafenauer Straße 22, 94078 Freyung
Germany
wolfgang.dorner@th-deg.de

ABSTRACT

The amount of historical photographs is a challenge for archives with regard to analyzing, documenting and structuring archival material and make it retrievable and accessible. Established human centric concepts will not be able to handle the instream of un- or semi-structured material or handle analogue documentation. We suggest the use of artificial intelligence to pre-classify and organize historic photographs in collections and archives. With regard to uncertainty the concept of artificial intelligence based classification is discussed based on two questions: (1) What is the uncertainty in current archival documentation, assuming that the quality and completeness of a current classification depends on the partial and restricted knowledge of a person? (2) Can algorithmic concepts help to quantify the level of uncertainty? The aspect of documentation is discussed based on a comparison of algorithms and classical processes. The contribution of artificial intelligence to quantify the uncertainty is tested in a case study. For this an artificial neural network / convolutional neural network was trained and tested on a collection of historical photographs.

CCS CONCEPTS

• **Computing methodologies** → Symbolic and algebraic manipulation → **Machine learning**;

KEYWORDS

Artificial intelligence, image classification, Tensorflow.

1 INTRODUCTION

The amount of historic photographs is challenging archives. Besides a continuous or increasing instream of new collections and objects also the digitisation and digital documentation of existing collections exceeds the capacities of most museums. The missing classification of archival material hinders the use of these photographs for scientific purposes or restricts the work either through incomplete or inconsistent documentation on already documented material or the restricted possibility to access undocumented material. Hence uncertainty arises first from the fact, that not all existing material can be used, because it is inaccessible, second the level of incompleteness or inconsistency can be quantified or only roughly estimated with an excellent knowledge of the whole documentation and is underlying classification strategies.

Established human centric concepts will neither be able to solve the problem of a growing instream, nor do they guarantee a consistent documentation. Human classification strongly depends on the knowledge and standardisation of classifications applied and provides a strong undefined bias with regard to the results.

We propose the use of artificial intelligence and machine learning to automatically analyse and classify historical photographs. With regard to uncertainty the concept of artificial intelligence based classification is discussed based on two questions: (1) What is the uncertainty in current archival documentation, assuming that the quality and completeness of a current classification depends on the partial and restricted knowledge of a person? (2) Can algorithmic concepts help to quantify the level of uncertainty?

The presented work hast three objectives: First to test the concept of convolutional neural networks and transfer learning for classification on a collection of historical photographs. Second to quantify the uncertainty in this automated classification approach, comparing different strategies and apply a trained neural network on a test data set. Third to discuss possibilities of applying this approach to larger collections and the resulting decrease of uncertainty of classification through continuous machine learning and automatic reclassification.

In a first step different uses of historical photographs and the underlying challenges for retrieval processes in archives are discussed based on a literature review. Resulting classification

concepts and requirements are discussed with regard to established concepts of artificial intelligence. Based on established best practices in image analysis a classification strategy and approach is proposed and tested on a small collection of historic photographs. Results are discussed with regard to the quantification of underlying uncertainty in the classification and an approach for the transfer on larger and more complex applications is suggested.

2 THEORY AND BACKGROUND

Classifications are an important concept for meta-data and retrieval of historic photographs in archival systems [2]. The discussion about expert classification versus social tagging shows the concerns with regard to quality of tags/classes as well as the problems resulting from personal influence of the tagging done by persons [5]. Uncertainty in the classification of archival photographic material can be defined as the accuracy of classification, which is the number of wrong classifications divided by the number of all classifications. With regard to the discussion in [5] we can conclude that accuracy in classification results from the knowledge of a person classifying the material in combination with the underlying and applied classification standard. The estimation of these uncertainties requires from a searching person a distinct knowledge of both aspects. Different classification standards will result in incomparable results and, hence, in problems to retrieve data. Individual persons and even the varying daily performance of one person results in a bias and problems to technically calculate the inaccuracy.

Since 2012 Deep Neural Networks (DNN) outperform alternative algorithms in the ImageNet Competition on Image Classification with regard to accuracy [1]. The competition results show that Inception V3 und especially V4 significantly surpassed alternative systems during the last years. Transfer Learning for DNNs is an interesting approach to train a neural network and apply this network on alternative but similar data sets. Lagunas & Garces applied this concept on the transfer from national images to artistic depictions [4]. With regard to historical photographs transfer learning could be an interesting strategy to train a DNN with classical images and start a transfer learning stage with small set of pre-classified historical photographs as training data. ImageNet provides a dataset of over 15 million images for training and testing. For these images almost 22,000 classes are available [3]. A training data set with manually added annotations as classification is available [6]. ANN for images might help to provide a stable and standardized classification for archival materials.

3 METHODOLOGY

The suggested concept was tested on a sample data set from Atelier Seidel, an archive and museum for historical photographs from the 1880ies till 1954. The archive is located in Český Krumlov, South Bohemia, and the majority of photographs stem from this region. From a total stock of 140,000 photographs 20,000 were digitized through past years and as set of roughly 2,000 will

be used in this study. For the purpose of this study a set of 629 was used for training and validation.

Artificial Neural Networks (ANN) depend on a set of pre-classified data to train the ANN. For the given case study, the set of available historical photographs as data is too small, to be split into a training, validation and a test data set and to use the training data set for machine learning. Hence, the concept of transfer learning was applied. We decided for Tensor Flow (TF) in the version 1.7.0 as the ANN Framework due to the existing tools, the ability to use CPUs as well as GPUs, the interfaces for individual programming languages for a later integration in a web application and results in image classification during past ImageNet Large-Scale Visual Recognition Challenges.

As a training data set for the basic training phase 1.2 mi. images from ImageNet [4] was used to train the Inception V3 and V4 ANNs based on TF. The restricted set of historical photographs from the sample provides the basis for transfer-learning.

A simplified classification of four classes was applied. Historical photographs were manually pre-classified and separated into portraits (120 photographs), landscape (254 photographs), group photographs (119 photographs) and buildings/architectural photography (136 photographs). Photographs can belong to multiple classes, but were pre-classified based on the class with the highest areal proportion of one class. Manually pre-classified photographs were used to train the ANN and later analysed the classification quality through comparing the result of the ANN with the manual pre-classification. Due to the strong generalisation of classes interferences became a problem e.g. for classes such as landscape and buildings, while most photographs with landscape also contain buildings.

Training happened on two stages: First the ImageNet data set was used to train the ANN based on classical photographs. Second transfer learning was applied on basis of the pretrained ANN and applying it to the small set of historical photographs. Datasets for both learning stages were split into training, validation and test data. Training and validation data was used in the training phase to train the ANN and validate the state of training based on the validation data set. After training the ANN was applied to the test dataset to test the final ability to classify and calculate the accuracy. We used 80% of data for training, 10%for validation and 10% for test.

All images from both datasets were standardized by size (299 x 299 pixels) and converted to greyscale. The ANN was trained using 200 and 400 training steps. For transfer learning the training batch size for each class was 50 photographs. The validation batch size was 64 and the test batch size 59 photographs.

Because an image can only belong to one class, the resulting error could be calculated by counting after the automated classification all images that have been assigned to the wrong class.

4   RESULTS

Inception v4 with 200 training steps achieved a final precision of 86,4% and with 400 steps a maximum of 88,1% (Fig. 1). In contrast Inception v3, with 200 steps had an accuracy of 100,0% and with 400 steps of 98,3%.

The development of classification accuracy during the validation already indicates that in general the V3 achieves in general a higher accuracy independent of the number of training steps (Fig. 2). During the tests v4 showed a higher accuracy during training but lower accuracy during validation, which may indicate an overfitting.
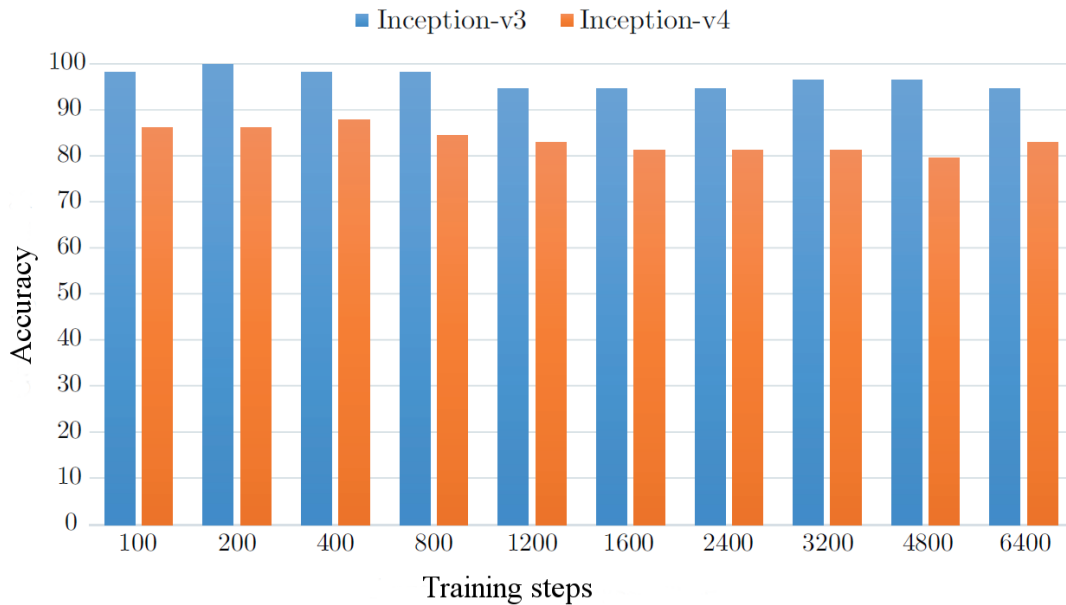

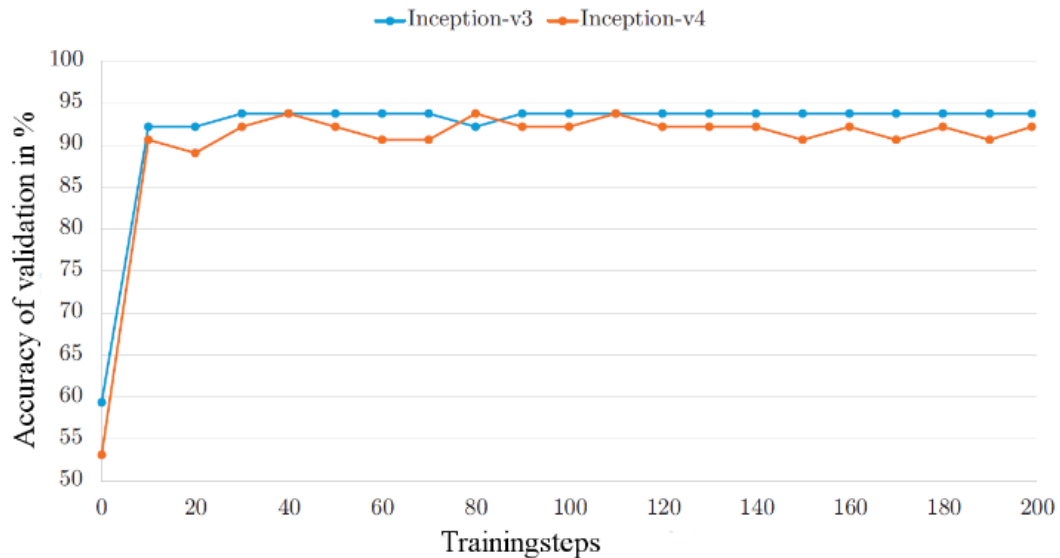**Figure 1. Training steps and accuracy of classification with the test dataset**

Figure 2. Accuracy of validation for 200 training steps

5 DISCUSSION

For the training of an ANN with historical photographs the insufficient number of pre-classified data is an obstacle. The process of transfer learning provided stable and accurate results even based on a small set of historic photographs. The ImageNet dataset showed to be a solid basis for the first training stage of transfer learning.

An explanation for the results with a low number of training steps (200-400) could be the small number of classes and the small training dataset. Further data indicates that the reduced accuracy of v4 in comparison to v3 may be the consequence of overfitting during training stage and needs further research and a larger set of photographs and images. A more specific choice of classes could help to reduce interferences but would as a consequence also require more photographs in the training phase.

The ANN showed a high accuracy. This accuracy will remain stable also for a repetition of the classification. Which means that ANNs would be able to reduce uncertainty in the classification of archival material due to standardized and stable classification scheme that can be applied also to larger amounts of data.

Through new training material the number of classes as well as the accuracy can be increased. And a newly trained ANN can be applied to already classified material only for the costs of the operation of the ANN. The stability of classification and the knowledge of the uncertainty in the classified material is a strong advantage over the human classification, that strongly relies on the individual general and daily performance and may vary over time and between persons. This means that ANN based classification can become a replicable and reliable method for classification tasks on historical photographs, given that a set of classified data is available for training.

To better analyse problems mentioned above and test the stability and classification accuracy for a more complex and diverse set of classes larger datasets will be necessary as well as different classification schemes.

6 CONCLUSION

The research was based on two leading questions: (1) What is the uncertainty in current archival documentation and (2) can algorithmic concepts help to quantify the level of uncertainty? The discussion of current classification concepts showed, that uncertainty mainly arises from the knowledge of a person classifying the material in combination with the underlying classification scheme applied as a standard.

The estimation of these uncertainties requires from a searching person a distinct knowledge of both aspects. Artificial intelligence can improve the understanding of uncertainty with regard to two dimensions.

First CNNs are able to provide a stable classification based on a calculable accuracy. Second, classification can become a recurring process, where new concepts of classification can be applied to already classified material on a stable basis. While training of the network results in a continuously improved classification quality, the speed of classification can be used to

continuously reclassify a collection based on the most up to date network only with computational costs. A continuously trained and improved network provides a standardized and developing basis for a classification, removing the bias of individual persons on the classification of parts of or between different collections.

In a first stage the neural networks was trained and applied to a small set of data. In the current development stage the CNN will be linked to an archival database of the project PhotoStruk (www.photostruk.de) and a recurring classification of the datasets in the database will be established for new classification schemes and newly trained neural networks.

The presented state of work is just a first attempt for a simplified pre-classification to analyse the potentials of transfer learning and its application to historical photographs.

In a next step first the full number of classes available in ImageNet data sets need to be applied and tested on historic photographs. Second the of concept of semiology and ontologies need to be integrated and tested on larger amounts of data.

The coupling of machine learning with human based classification, especially citizen science and crowd sourcing, to collect new information to train an ANN, could be interesting approaches to further develop classifications concepts.

The web platform developed as part of the project PhotoStruk includes elements for citizen science and volunteered geographic information. In a future stage data from user tagging and commenting could provide a new trainings dataset for transfer learning and increase and widen the classification scheme.

REFERENCES

[1] Canziani, A., Paszke, A. and Culurciello, E. 2016. An Analysis of Deep Neural Network Models for Practical Applications. *arXiv:1605.07678 [cs].* (May 2016)

[2] Choi, Y. and Rasmussen, E.M. 2003. Searching for images: The analysis of users' queries for image retrieval in American history. *Journal of the American Society for Information Science and Technology.* 54, 6 (Apr. 2003), 498–511.

[3] Krizhevsky, A., Sutskever, I. and Hinton, G.E. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM.* 60, 6 (May 2017), 84–90.

[4] Lagunas, M. and Garces, E. 2018. Transfer Learning for Illustration Classification. *arXiv:1806.02682 [cs, stat].* (May 2018). DOI:https://doi.org/10.2312/ceig.20171213.

[5] Lu, C., Park, J. and Hu, X. 2010. User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. *Journal of Information Science.* 36, 6 (Jan. 2010), 763–779. DOI:https://doi.org/10.1177/0165551510386173.

[6] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. and Fei-Fei, L. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision.* 115, 3 (2015), 211–252.