Architectural Style Classification for Mobile Applications

Project proposal for the course Applied Deep Learning at TU Wien

Maximilian Hannes Reinhardt

24 October 2023

1 Introduction

This document proposes the development of an architectural style classification model that is eligible for the use on mobile devices as a project for the course Applied Deep Learning at TU Wien. The project aims at developing a deep learning model that requires low computational resources in performing inference while achieving adequate accuracy. The intended approach combines the established MobileNets architecture [1] with an attention mechanism that is also used in the architectural style classification model proposed by Wang et. al. in 2023 [3] which can be considered as state-of-the-art. For the training and testing of the model, an established dataset for the classification of architectural styles is used, so that a benchmarking to other models is possible. The project topic would therefore be fine-grained image classification. The project type can be classified as "bring your own method" within the given categories.

The next sections summarize the proposed project by specifying the project idea, modelling approach, dataset and work-breakdown. At the end of the document, key resources are referenced.

2 Project Idea and Motivation

Architectural style classification has a number of possible applications such as tourism, the analysis of historical buildings, or architectural education [3]. Moreover, it is an interesting task from a scientific point of view, since architectural styles can exhibit strong similarities and since some buildings show elements of different styles [3, 5]. In the past, models were mainly published that aim at achieving the highest possible accuracy. Especially since the emergence of deep learning techniques, the models have become very large so that predictions with the models require a high computational effort.

However, especially in the field of tourism, the use of architectural style classification models on mobile devices is interesting. Such an app could be an exciting tool for architecture-enthusiastic users to get to know new cities. For this purpose, a model that can perform inference with low computational effort is particularly suitable. This would also be beneficial from a sustainability perspective, if a larger number of people would use such a model to learn about cities.

Therefore, the goal of this project is to create a deep learning model for the classification of architectural styles that causes low computational effort in the generation of predictions and still achieves a reasonable accuracy.

3 Approach

The approach proposed for this project work combines the MobileNets architecture [1] with the Convolutional Block Attention Module (CBAM) developed by Woo et. al. [4]. This follows the idea from Wang et. al. [3] to use CBAM for the extraction of spatial features in architectural style classification.

The model proposed by Wang et. al. in 2023 [3], which at least in some settings is currently the best performing model, consists of four parts. After a selective preprocessing module, features are extracted using the Inception-v3 model. The output of this module is then processed with CBAM before it is passed to a softmax classifier. The CBAM consists of a channel attention module and a spatial attention module [4]. Thus, it helps the model to focus on the most important features and to incorporate spatial information. In the case of a building that combines different architectural styles, for example, the positions of the different style elements might be important for classification [3].

The established MobileNets offer an efficient convolutional neural network (CNN) architecture that is based on depth-wise separable convolutions [1]. Howard et. al. [1] recommend to adjust the size of the network by changing the resolution of the input images and uniformly shrinking the number of channels at each layer.

The model proposed in this project follows the MobileNets architecture but further incorporates a CBAM in the last layers corresponding to the approach of Wang et. al. [3]. It is motivated by the facts that MobileNets have proven to be effective in a lot of image processing tasks [1] and CBAM appeared to be eligible to enhance the ability of a generic CNN feature extractor to classify architectural styles [3]. At the same time, CBAM has a low impact on the number of parameters [4] such that the efficiency of MobileNets should be retained.

4 Dataset

For the training and testing of the proposed model the architectural style dataset published by Xu et. al. [5] in 2014 is used. The dataset contains colored images of buildings of 25 architectural styles. Each class has 60 to 300 associated instances summing up to approximately 5000 images in total.

Since the release, the dataset has been used for a number of publications, including the model of Wang et. al. [3] on which the approach proposed in this project is based, so that benchmarking is possible.

As the original distribution of the dataset is no longer available, the data is obtained from a re-upload on Kaggle [2].

5 Work-breakdown

According to the course requirements, the project includes the steps dataset collection, model building, model training and fine-tuning, application development, writing the final report, preparing the presentation. In this project, it is planned to perform model building and model training iteratively by re-implementing a standard MobileNet at first and then extending the model by integrating the CBAM. Fine-tuning will be conducted for the final model. Table 1 shows the single steps and the associated effort in days. In total, the upcoming tasks will take approximately 16 days. One day is intended to correspond to approximately eight hours.

Task	Effort (in days)
Dataset collection	0
Implementing MobileNet	2.5
Training MobileNet	2.5
Integrating CBAM	2
Training MobileNet + CBAM	1
Fine-tuning	1
Application development	3
Report writing	3
Presentation preperation	1

Table 1: Tasks and associated effort in days. One day is intended to correspond to approximately eight hours.

References

[1] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint* arXiv:1704.04861.

[2] Kaggle. (n.d.). *Architecture dataset.* https://www.kaggle.com/datasets/wwymak/architecture-dataset

[3] Wang, B., Zhang, S., Zhang, J., and Cai, Z. (2023). Architectural style classification based on CNN and channel–spatial attention. *Signal, Image and Video Processing, 17*(1), 99-107.

[4] Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. *Proceedings of the European conference on computer vision (ECCV).*

[5] Xu, Z., Tao, D., Zhang, Y., Wu, J., Tsoi, A.C. (2014) Architectural style classification using multinomial latent logistic regression. *Proceedings of the European conference on computer vision (ECCV).*