氏名	TAN WEI REN
学位の種類	博士（工学）
学位記番号	甲　第　687　号
学位授与の日付	平成 30 年 3 月 20 日
学位授与の要件	信州大学学位規程第 5 条第 1 項該当
学位論文題目	Deep learning for artworks analysis and synthesis （アートワーク分析合成のための深層学習）
論文審査委員	主査　教　授　田中　清 准教授　Aguirre　Hernan 教　授　丸山　稔 准教授　白井　啓一郎 准教授　児玉　明　（広島大学）

論　文　内　容　の　要　旨

With the rapid technology advancement, paintings digitalization has been made possible. Hence, it becomes desired to solve various artwork related problems through computer with the assistance from machine learning-based solutions. Some interesting and important applications include archiving digital artworks, painting analysis, image stylization, etc. However, research in this domain has received limited attention, mainly due to the difficulty in collecting the artworks to create a proper artworks dataset. Fortunately, many digital fine-art paintings have been made available online over the past few years, making the access to these digital artworks easier than ever. Recently, a large scale Wikiart dataset was introduced and released, encouraging more research in this domain.

Among the available machine learning techniques, deep learning has received much attention in the last decades. Evidently, deep learning is currently the state-of-the-art solution to many computer vision problems. More interestingly, deep learning enables automatic features learning, contrast to the traditional methods where the features are human engineered. However, most works are only focusing on datasets with more structural subjects, e.g. digits, faces, and natural objects. Compared to these works, recognition of artworks is more challenging because a lot of art categories require understanding of non-structural cues, e.g. perspective, emotion, history, etc. In addition, many artworks are non-figurative or structured abstract.

Overall, this thesis is interested in the application of deep learning on artworks. In particular, this thesis aim to *understand the features learning of artworks using deep models* because features extraction is the first step to many computer vision problems, including artworks analysis. For this purpose, deep models are trained for artworks *classification* and *synthesis* tasks and analyzed via various visualization methods. This will reveal what kind of features can the deep models learn from the artworks. Furthermore, this thesis also concerned on the *issues arise when embedding deep models into real-world devices*. To this end, this thesis is divided into two parts.

The first part of this thesis first focuses on training and understanding deep *discriminative* models for artworks classification tasks, i.e. categorizing the *genre*, *artist*, or *style* of an artwork. In this work, a modified Wikiart dataset is used and released to encourage better comparative studies in future works. In particular, the dataset is explicitly split into training and validation sets. The best models found employed and finetuned an ImageNet pre-trained model, achieving state-of-the-art results. Then, the neurons' activations are visualized in this work to analyze the features learned. The visualizations show that the models are able to learn the structural visual cues from the artworks. Meanwhile, it is also found that the differences between the art categories can be extremely abstract and hard to recognize. For instance, the difference between Baroque and Rococo lies on the emotions felt from the paintings. Grasping such abstract concepts is difficult even for a human, while it is unclear that if the models are able to capture such cues with the visualization technique used.

Next, a Generative Adversarial Network (GAN) variant, named **ArtGAN** is proposed and trained on the Wikiart dataset conditioned on the given *genre*, *artist*, or *style*. GAN is a type of *generative* model that learns to simulate true data from the joint distribution, hence able to learn richer representations compared to discriminative model. More importantly, GAN is able to synthesize photo-realistic images compared to other generative models, which is extremely useful when assessing the features learned by the models. The contributions of the proposed ArtGAN are three-fold: 1) The labels are leveraged when calculating the loss function to improve the image quality for each category; 2) Autoencoder is incorporated to compute energy-based loss function for additional complementary information; 3) Most importantly, a novel **magnified learning** is proposed to learn the correlations between neighbouring pixels better, improving image quality. In addition, it allows synthesizing images at higher resolution compared to the maximum resolution available in the dataset. For quantitative assessment, the proposed ArtGAN is trained on CIFAR-10 and STL-10 datasets, and achieves state-of-the-art results on Inception score. Meanwhile, the results show that the proposed ArtGAN is able to generate high resolution photo-realistic images. Furthermore, the synthesized artworks show that deep models seem able to learn some abstract characteristics from the artworks through visual cues.

In the second part of the thesis, the applicability of deep models in real-world devices is explored. In practice, it is infeasible to embed deep models into resource limited hardware, e.g. mobile devices. This is because many deep models are designed to have extremely high memory requirement. To address this problem, it is desired to reduce the memory requirement with minimum compensation on the model performance. In this work, a novel one-shot deep compression method based on the fuzzy quantity space is proposed to remove redundant weights. The experiments demonstrate that the proposed approach is able to compress the deep models up to 14 times with a minimal loss of classification accuracy.

This thesis is concluded with a summary of its contributions. Last but not least, promising directions of future works are outlined.