摘要

在信息时代，图像与人们的生活密不可分。图像常常会出现失真，使其视觉质量降低，图像质量评价这一问题因此产生。图像质量评价方法首先可分为主观和客观两种，主观方法可靠但费时费力，因此需要高效、快捷的客观模型。客观模型根据对参考图像信息的利用程度可分为全参考、部分参考和无参考三种，其中无参考模型实用价值更高也更困难，是本文主要的研究对象。客观无参考图像质量评价模型的发展经历了由传统的人工提取特征，到使用深度神经网络自动提取特征的过程。考虑图像的失真类型，图像质量评价又可以分为合成失真和真实失真两种应用场景，其中真实失真场景更复杂也更有实用价值。无参考图像质量评价问题目前面临着两个挑战：一方面由于主观评价的高成本导致有标签图像质量评价数据集普遍规模较小，容易使深度网络过拟合；另一方面真实失真场景的图像质量评价问题逐渐受到关注，该问题比合成失真更加困难，很多适用于合成失真的模型在真实失真上表现不好。对此，本文关注利用无标签数据提取质量特征，以帮助合成失真与真实失真场景的质量评价的问题，并从模型构建和模型应用两个方面对问题进行研究。模型构建方面，我们理清了图像的质量、失真与内容这几个概念的含义与关系，根据图像失真与内容间的独立性和互补性提出了基于图像重建任务的自监督学习框架。该框架在无标签数据集上预训练，能够提取图像的失真特征并用于图像质量评价。模型应用方面，本文分别研究了模型在合成失真与真实失真场景的应用。下面进一步介绍这两个方面。

模型构建：图像质量取决于人类的视觉感知，这一概念较为主观。从图像作为信息载体的本质来看，失真描述信息传递过程是否受到损坏，内容描述图像传递的信息，这两个概念更加客观。对于三者的关系，失真与内容具有一定的独立性和互补性，失真是影响图像质量最重要的因素，但图像质量也受内容的影响，这一点本文不作深入研究。根据图像失真与内容的独立性和互补性本文考虑使用图像重建这一自监督学习任务，在已知内容特征的前提下提取图像的失真特征，以帮助图像质量评价任务。内容特征在框架中起着重要的作用，其不能包含失真信息，以保证失真特征的提取。

模型应用：我们首先将自监督框架应用于合成失真场景，由参考图像提供内容特征。实验发现模型提取的失真特征有着良好的质量预测能力和可视化效果。同时我们发现模型只能对见过的失真类型有良好的特征提取能力，该问题可以通过扩充预训练阶段的失真类型来解决。真实失真场景由于缺少参考图像，内容特征只能由失真图像提供，但其中可能包含失真信息。为解决这一问题，本文发现通过构建失真图像域和清晰图像域能够区分某一图像特征是否与失真相关，并使用特征在两个图像域上分布的

标准差之比对判别方式进行量化，之后设计实验对其进行了验证。我们将这一特征的判别方式应用于内容特征通道的筛选，以起到屏蔽内容特征中的失真信息的目的，从而使自监督框架能够应用于真实失真场景。实验发现真实场景下提取的特征有一定的质量预测能力。

关键词：无参考图像质量评价，自监督学习，失真特征，合成失真，真实失真

No-Reference Image Quality Assessment Based on Self-Supervised Learning

Yingjian Song (Computer Applied Technology)
Directed by Prof. Tingting Jiang

ABSTRACT

In the information age, images are inseparable from people's lives. Images often suffer from distortions, degrading their perceptual quality, and the problem of image quality assessment arises. The problem can be first divided into subjective and objective. Subjective methods are reliable but time-consuming and laborious, so efficient and fast objective models are needed. According to utilization degree of the reference image information, objective models can be divided into three categories: full-reference models, reduced-reference models and no-reference models. Among them, no-reference models have higher practical value and are more difficult, which are the main research object of this dissertation. The development of objective no-reference image quality assessment models has experienced a process from traditional manual feature extraction to automatic feature extraction using deep neural networks. Considering the image distortion, image quality assessment can be divided into two application scenarios: synthetic distortions and authentic distortions. The authentic distortion scenario is more complex and more practical. The problem of no-reference image quality assessment currently faces two challenges. On the one hand, due to the high cost of subjective methods, the labeled image quality assessment datasets are generally small in size, which is easy to make deep neural networks overfitting. On the other hand, the problem of image quality assessment for authentic distortion scenes has gradually attracted attention. This problem is more difficult than synthetic distortion scenes, and many models suitable for synthetic distortions do not perform well on authentic distortions. In this regard, we focuses on the problem of using unlabeled data to extract quality features to help the image quality assessment of synthetic distortions and authentic distortions, and study the problem from two aspects of model construction and model application. In terms of model construction, we clarify the meaning and relationship of the concepts of image quality, image distortion and image content, and propose a self-supervised learning framework based on the image reconstruction task which is inspired

from the complementarity between image distortion and image content. The framework is pre-trained on unlabeled dataset and is able to extract distortion features of images and used them for image quality assessment. In terms of model application, we study the application of the model in the synthetic distortion and authentic distortion scenes respectively. These two aspects are further described below.

Model construction: Image quality depends on the human visual perception, which is subjective. From the essence of image as the information carrier, distortion describes whether the information transmission process is damaged, and content describes the information transmitted by the image. These two are more objective concepts. Distortion and content are independent and complementary. Distortion is the most important factor affecting the quality, but image quality is also affected by its content, which will not be further discussed in this paper. According to the independence and complementarity of distortion and content, we consider using image reconstruction as a self-supervised learning task to extract the distortion features of the image under the premise of knowing the content features. Content features play an important role in this framework, which cannot contain distortion information to guarantee the extraction of distortion features. We hope that the model extracts high-quality distortion features to help the image quality assessment problem, regardless of the relationship between quality and content.

Model application: We first apply the self-supervised framework to the synthetic distortion scenes, with the content features provided by the reference images. Experiments show that the distortion features extracted by the model have good quality prediction ability and visualization effect. At the same time, we find that the model can only have good feature extraction ability for seen distortion types, and this problem can be solved by augmentating the distortion types in the pre-training stage. Due to the lack of reference images for authentic distortion scenes, the self-supervised framework cannot be directly applied. To solve this problem, we find that it is possible to distinguish whether an image feature is related to image distortion by constructing image domains, and quantify the discriminant method by using the ratio of the standard deviation of the feature distribution in the two image domains. And then we design experiments to verify it. We apply this discriminative way of features to the filtering of the content feature channels, in order to mask the quality information in the content features, so that the self-supervised framework can be applied to the authentic distortion scenes. The experiments have found that the features extracted in the authentic distortion scenes have certain quality prediction ability.

KEYWORDS: No-reference image quality assessment, Self-supervised learning, Distortion feature, Synthetic distortion, Authentic distortion