# Using Deep Learning to automatically rank millions of hotel images 使用深度学习自动排列数百万酒店图像

At (the leading price comparison website in Europe and one of the largest portals in the German e-commerce market) we provide one of the best hotel price comparisons available on the market. For each hotel we receive dozens of images and face the challenge of choosing the most “attractive” image for each offer on our offer comparison pages, . Given that we have millions of hotel offers, we end up with more than 100 million images for which we need an “attractiveness” assessment.  
在（欧洲领先的价格比较网站和德国电子商务市场最大的门户网站之一）我们提供市场上最好的酒店价格比较之一。对于每家酒店，我们都会收到几十张图片，并面临在我们的报价比较页上为每个报价选择最具“吸引力”的图片的挑战。考虑到我们有数百万的酒店优惠，我们最终得到了超过1亿张图片，我们需要对这些图片进行“吸引力”评估。

We addressed the need to automatically assess image quality by implementing an aesthetic and technical image quality classifier based on Google’s research paper “”. NIMA consists of two Convolutional Neural Networks (CNN) that aim to predict the aesthetic and technical quality of images, respectively. The models are trained via transfer learning, where pre-trained CNNs are fine-tuned for each quality classification tasks.  
我们通过实现基于谷歌研究论文“”的美学和技术图像质量分类器来解决自动评估图像质量的需要。NIMA由两个卷积神经网络（CNN）组成，分别用于预测图像的美学和技术质量。这些模型通过转移学习进行训练，在转移学习中，预先训练的cnn针对每个质量分类任务进行微调。

In this article, we will present our training approach and insights that we’ve gained throughout the process. We will then try to shed some light on what the trained models actually learned by visualising the convolutional filter weights and output nodes of our trained models.  
在本文中，我们将介绍我们在整个过程中获得的培训方法和见解。然后，我们将试图通过可视化训练模型的卷积滤波器权重和输出节点，来阐明训练模型实际学到的知识。

We’ve published the trained models and code on . The provided code allows one to use any of the pre-trained CNNs in , so we are looking forward to contributions that explore other CNNs for image quality assessments 😃  
我们已经发布了训练模型和代码。所提供的代码允许用户在中使用任何经过预训练的CNN，因此我们期待有助于探索用于图像质量评估的其他CNN

### Training 培训

The aesthetic and technical classifiers were trained in a transfer learning setup. We used the with ImageNet weights, and replaced the last dense layer in MobileNet with a dense layer that outputs to 10 classes (scores 1 to 10).  
审美和技术分类器是在迁移学习环境中训练的。我们使用WITHIMAGENET权重，并将MobileNet中的最后一个密集层替换为输出到10个类（得分1到10）的密集层。

#### Earth Mover’s Loss 土方机械损失

A special feature of NIMA is the use of the Earth Mover’s Loss (EML) as the loss function, contrary to the Categorical Cross Entropy (CCE) loss, that is generally applied in Deep Learning classification tasks. The EML can be understood as the amount of “earth” that needs to be moved to make two probability distributions equal. A useful attribute of this loss function is that it captures the inherent order of the classes. For our image quality ratings, the scores 4, 5, and 6 are more related than 1, 5, and 10, i.e. we would like to punish a prediction of 4 more if the true score is 10 than when the true score is 5. CCE does not capture this relationship, and it is often not required in object classifications task (e.g. misclassifying a tree as a dog is as bad as classifying it as a cat).  
NIMA的一个特点是使用地球运动损失（EML）作为损失函数，而不是一般应用于深度学习分类任务的分类交叉熵（CCE）损失。EML可以理解为需要移动以使两个概率分布相等的“地球”的数量。这个loss函数的一个有用属性是它捕获类的固有顺序。对于我们的图像质量评级，分数4、5和6的相关性大于1、5和10，即如果真分数为10，我们希望惩罚预测值4，而不是真分数为5。CCE不捕获这种关系，在对象分类任务中通常不需要它（例如，将树误分类为狗和将树分类为猫一样糟糕）。

In order to use the EML we need for each image a distribution of ratings across all ten score classes. For the dataset, which is used to train the aesthetic classifications, these distribution labels are available. For the dataset, used for the technical classifications, we inferred the distribution from the mean score given for each image. For more details on our distribution inference check out our .  
为了使用EML，我们需要为每个图像在所有十个评分类中分配评分。对于用于训练美学分类的数据集，可以使用这些分布标签。对于用于技术分类的数据集，我们根据每个图像的平均得分推断分布。有关我们的分布推断的更多详细信息，请参阅。

#### Fine-tuning stages 微调阶段

We train the models in a two stage process:  
我们分两个阶段对模型进行培训：

1. We start by training only the last dense layer with a higher learning rate to ensure that the newly added random weights are adjusted to the ImageNet convolutional weights. Without this burn-in period you risk juggling around the convolutional weights at training start and consequently slowing down the training process.  
   我们首先只训练最后一个具有较高学习率的稠密层，以确保新添加的随机权重调整为ImageNet卷积权重。如果没有这段磨合期，你就有可能在训练开始时绕着卷曲的重物转来转去，从而减缓训练过程。
2. After the burn-in period we train all weights in the CNN with a low learning rate.  
   在磨合期之后，我们在CNN中以低学习率训练所有的重量。

For both the aesthetic and technical model the train and validation losses level out after 5 and 25 epochs, respectively. This is a good indicator that the newly added weights have learned to classify aesthetics and technical quality as good as possible, and it is time to start training all weights.  
对于美学和技术模型，列车和验证损失分别在5个和25个阶段后趋于平衡。这是一个很好的指标，表明新增加的权重已经学会了将美学和技术质量尽可能好地分类，现在是开始训练所有权重的时候了。

For the aesthetic classifier we see a significant drop in loss once we start training also the convolutional weights (dashed lines in left graph above), indicating that we are adjusting the convolutional weights quite a bit for the aesthetic classification task. For the technical classifier the drop in loss is smaller, which at first is counter-intuitive, as the technical image quality should be object agnostic, and the ImageNet weights are optimised to recognise objects. The small drop might be due to the very small learning rate that is required to regularise training on the small TID2013 dataset.  
对于美学分类器，一旦我们开始训练卷积权重（上图中的虚线），损失就会显著下降，这表明我们正在为美学分类任务调整卷积权重。对于技术分类器，损失下降较小，这首先是违反直觉的，因为技术图像质量应该是对象不可知的，并且图像净重被优化以识别对象。这一小幅下降可能是由于在小型TID2013数据集上进行正规化培训所需的学习率非常低。

You can find all hyper-parameters used for training on our .  
你可以在我们的上找到所有用于训练的超参数。

### Results 结果

The above predictions show that the aesthetic classifier correctly ranks the images from very aesthetic (leftmost image with sunset) to least aesthetic (boring hotel room on the right). Similarly for technical classifications, the classifier predicts higher scores for undistorted images (first and fourth image from left), versus images with jpeg compression (second and fifth) or blur (third and sixth).  
上述预测表明，审美分类器正确地将图像从非常唯美（最左边的图像是日落）到最不唯美（右边是无聊的酒店房间）。与技术分类类似，分类器预测未失真图像（左一和左四）的分数更高，而jpeg压缩图像（第二和第五）或模糊图像（第三和第六）。

### Visualisations 形象化

In order to gain a better understanding as to how the CNN assesses aesthetic image quality, we used the package to visualise the learned convolutional filter weights and output nodes in Aesthetic MobileNet. The awesome blog post provides a great interactive overview of state-of-the-art CNN visualisation techniques.  
为了更好地了解CNN是如何评估美学图像质量的，我们使用该软件包来可视化美学MobileNet中学习到的卷积滤波器权重和输出节点。这篇很棒的博客文章提供了最先进的CNN可视化技术的交互式概述。

Earlier convolutional layers are generally associated with simpler structures, like edges, wave patterns, and grids. The images above show patterns associated to six filters in layer 23 of MobileNet - the six images in the top row are generated from the original MobileNet ImageNet weights (ImageNet MobileNet), whereas the bottom row images are generated from the MobileNet weights fine-tuned on the AVA dataset for aesthetic ratings (Aesthetic MobileNet). From the filter visualisations we can see that the earlier convolutional filters are not much affected throughout fine-tuning, as they are very similar to the original ones.  
早期的卷积层通常与更简单的结构相关联，如边缘、波型和网格。上面的图像显示了与MobileNet的第23层中的六个过滤器相关联的模式-顶行中的六个图像是从原始MobileNet ImageNet weights（ImageNet MobileNet）生成的，而底行图像是从在AVA数据集上微调的用于美学评级（美学MobileNet）的MobileNet weights生成的。从滤波器的可视化可以看出，早期的卷积滤波器在整个微调过程中没有受到太大的影响，因为它们与原始滤波器非常相似。

For mid convolutional filters at layer 51, the learned shapes are more complex, and resemble interwoven structures like fur or a grid with eyes. Even at this level, Aesthetic MobileNet is very similar to ImageNet MobileNet.  
对于第51层的中卷积滤波器，学习到的形状更为复杂，类似于毛发或带眼睛的网格等交织结构。即使在这个层次上，美感MobileNet与ImageNet MobileNet非常相似。

The later convolutional layers show even more complex structures that resemble animals and tree like shapes. We can see that the filters for Aesthetic MobileNet differ significantly from the ImageNet ones, as they seem to be less focussed on objects, e.g. no animal shapes in the fourth filter from the left.  
后面的卷积层显示了更复杂的结构，类似于动物和树的形状。我们可以看到，美感MobileNet的滤镜与ImageNet的滤镜明显不同，因为它们似乎不太关注对象，例如，在左边的第四个滤镜中没有动物形状。

We also generated visualisations for the output nodes of Aesthetic MobileNet , which represent the probabilities for scores 1 to 10. The visualisations thus show a “representative” image that is associated to each score.  
我们还为美感MobileNet的输出节点生成了可视化效果，表示得分1到10的概率。因此，可视化显示与每个分数相关联的“代表性”图像。

It is difficult to interpret the output node visualisations, just as much as it is difficult to define aesthetics. If anything, the visualisations for lower scores seem to be less colourful and diversified, whereas higher scores are associated with more colourful and dramatic shapes. The image for score 10 seems to resemble a landscape with a sky background, a motive generally associated with high aesthetics.  
很难解释输出节点的可视化，就像很难定义美学一样。如果有什么不同的话，低分数的视觉效果看起来就不那么丰富多彩了，而高分数的视觉效果则与色彩更丰富、更具戏剧性的形状联系在一起。得分10的图像看起来像是一幅天空背景的风景画，这一动机通常与高度的美学有关。

### Summary 摘要

In this article, we presented our business challenge to automatically assess the quality of images. We showed that the trained aesthetic and technical models successfully rank images according to aesthetics and technical quality. We further explored the learned CNN weights of the aesthetic model by visualising the convolutional filters and output nodes, and concluded that fine-tuning primarily affects later convolutional weights.  
在本文中，我们提出了自动评估图像质量的业务挑战。结果表明，经过训练的审美模型和技术模型成功地根据审美和技术质量对图像进行了排序。通过对卷积滤波器和输出节点的可视化，我们进一步探索了美学模型的学习CNN权重，并得出结论：微调主要影响后期卷积权重。

Fine-tuning deep neural networks is a great strategy to tackle many computer vision problems that businesses face. However, the classifications of these models, with their millions of parameters, are generally difficult to interpret, and we hope to have shed some light on this black box with our visualisation analysis.  
微调深层神经网络是解决企业面临的许多计算机视觉问题的一个伟大策略。然而，这些模型的分类，以及它们的数百万个参数，通常很难解释，我们希望通过我们的可视化分析，能够对这个黑匣子有所启发。

Please let me know if you found this article useful (👏🏻) so others can find it too, and share it with your friends. You can follow me here on Medium () or on Twitter ([@chris\_lennan](https://twitter.com/chris\_lennan)) to stay up-to-date with my work. Thanks a lot for reading!  
如果你觉得这篇文章有用，请告诉我，这样其他人也可以找到它，并与你的朋友分享。您可以在Medium（）或Twitter（[@chris\_lennan]（https://Twitter.com/chris\_lennan））上关注我，了解我的最新工作。非常感谢你的阅读！