第十三次作业

张康杰 10215102471

一、实验内容

1、 阅读 1 篇相关文献 撰写阅读报告
2、 下载彩色化处理的算法代码 尝试阅读代码 并运行 。（选做题）
①解读关键代码
②查看实验结果

二、Colorful Image Colorization

1、作者团队

作者团队是来自加州伯克利的Richard Zhang, Phillip Isola和Alexei A. Efros

2、研究问题

过往的图像上色问题要么依赖于与用户的交互要么存在去饱和的问题，而本文提出了一种基于CNN结构的自动的图像上色方法，能够超越过去方法的表现。并且，本文还发现，colorization是一个非常好的pretext task，作为一个cross-channel encoder能够很好地帮助自监督的训练。本文提出的上色方法在各个benchmark上都取得了SOTA的表现。

3、核心思想

对于一张灰度图，作者发现场景的语义信息(semantics)和表面纹理(surface texture)能够提供足够多的线索，比如天空是蓝色的，草地是绿色的等等。尽管有的时候这些semantic prior并不是对于所有的事物都奏效(有的时候物体真实的颜色并不与预测的相同，可能会有很多种情况)，但这不是本文关注的重点，本文的重点在于产生一张合理的彩色图片，足以使人辨别不出是由机器产生的即可。

通过输入明度图像L，模型会预测CIELAB色彩空间中的a，b色彩通道。（CIELAB色彩空间由明度通道L，红色到绿色的分量a和蓝色到黄色的分量b三个通道构成）。

任何的彩色图像都可以作为训练数据，而且不需要ground truth，只需要将其的ab通道作为监督信号，就可以通过自监督的方式训练。

先前的利用CNN来上色的一些方法会造成去饱和。这是因为其训练使用的loss function是从回归问题中借鉴过来的，是一个prediction和ground truth之间的欧式距离的一个loss function，训练过程就是最小化该loss function。但是这种loss function并不能很好地满足需求。因此作者就采用了另一种loss。先前的论文指出，color prediction是具有multi-modal的特性的，一个物体可以由多种可能的颜色，因此为了解决这个问题，本文中对于每一个pixel，预测其distribution of possible colors，并且同时更改了loss中的各项权重，用以突出rare color。

最后根据distribution of possible colors，将这个分布的annealed-mean作为最后的颜色。

4、解决方案

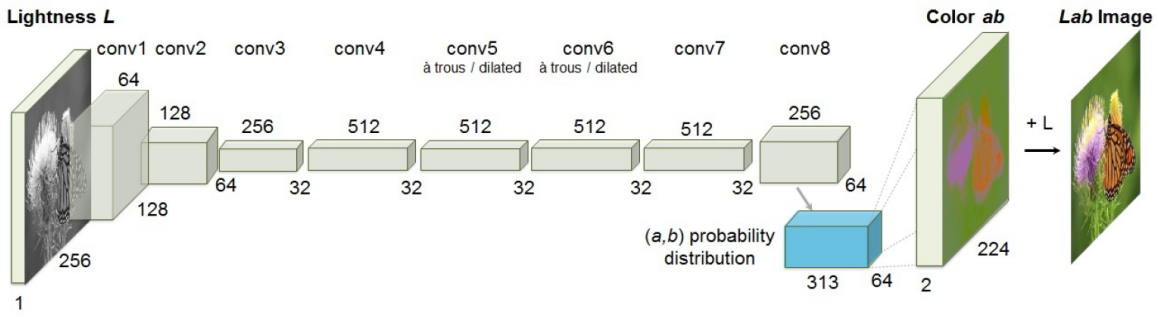


Fig. 2. Our network architecture. Each `conv` layer refers to a block of 2 or 3 repeated `conv` and `ReLU` layers, followed by a `BatchNorm` [30] layer. The net has no `pool` layers. All changes in resolution are achieved through spatial downsampling or upsampling between `conv` blocks.

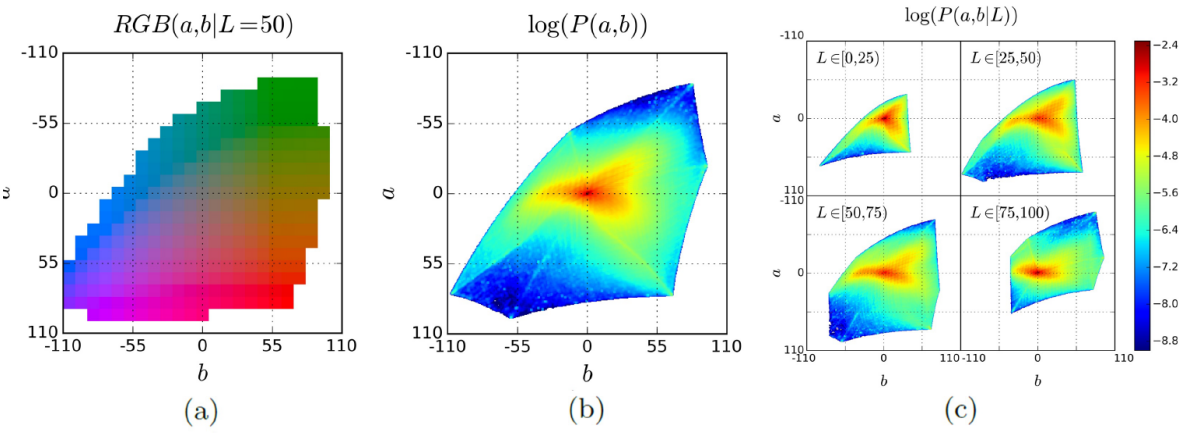
本文中通过训练一个CNN来实现从一张灰度图map到一个color distribution。

输入明度图像 $\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$，通过学习一个mapping $\widehat{\mathbf{Y}} = \mathcal{F}(\mathbf{X})$，来输出另外两个色彩通道的图像$\mathbf{Y} \in \mathbb{R}^{H \times W \times 2}$。以上的过程都是进行在CIELab空间中的，这是因为在CIELab空间中的distance就是perceptual distance。

$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

但是采取L2 loss存在着以下问题：

L2 loss实际上会造成averaging effect，这会导致欠饱和的问题。并且上色这一任务具有multimodal的特性，即同一个像素有多种合理的色彩，L2 loss并不能很好地解决这个问题。



因此作者团队把上述的问题看作是一个多分类问题(multinomial classication)。本文将ab空间分割成大小为10的grids，并且旨在色域空间内保存Q = 313个值，这样将色域空间离散化处理。然后学习一个mapping $\widehat{\mathbf{Z}} = \mathcal{G}(\mathbf{X})$将输入的lightness X转换成颜色的概率分布Z-hat，$\widehat{\mathbf{Z}} \in [0,1]^{H \times W \times Q}$

而为了优化这一个mapping $\mathbf{g}$，需要将Z与ground truth相比较。但是ground truth并不是以color probability distribution形式存在的，因此需要转换。这里定义了一个函数 $\mathbf{Z} = \mathcal{H}_{gt}^{-1}(\mathbf{Y})$来将ground truth转换成Z。

通过最小化Z-hat和Z的multinomial cross entropy loss优化$\mathbf{g}$。

$$\mathrm{L}_{cl}(\widehat{\mathbf{Z}}, \mathbf{Z}) = -\sum_{h,w} v(\mathbf{Z}_{h,w}) \sum_{q} \mathbf{Z}_{h,w,q} \log(\widehat{\mathbf{Z}}_{h,w,q})$$

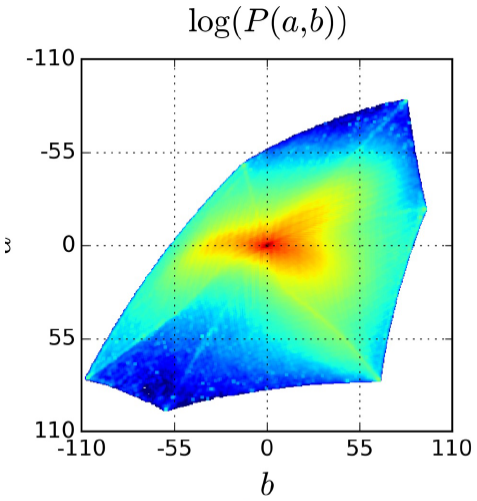
其中**v**是权重项，用于根据颜色的稀有度来赋予适当的权重，以平衡loss function。

最后，为了从color probability distribution转换到最终的pixel color的结果，通过mapping function **H** 来实现$\widehat{\mathbf{Y}} = \mathcal{H}(\widehat{\mathbf{Z}})$

Class rebalancing

自然图像中，颜色分量ab的值会严重地偏向更低的值。这是由于背景区域的颜色造成的。在ImageNet上的统计结果也表明这一现象。

如下图所示：



(b)

如果不考虑到这一点的话，loss function将仅仅受欠饱和的色彩值(即ab值很低)影响，受饱和色彩(即ab值很高)的影响非常小。最后的生成的彩色图片还是会出现欠饱和的现象。因此本文中对于上述的loss function加了一个权重来平衡。

权重的计算如下式。

$$v(\mathbf{Z}_{h,w}) = \mathbf{w}_{q^*}, \text{ where } q^* = \arg\max_q \mathbf{Z}_{h,w,q}$$

$$\mathbf{w} \propto \left((1-\lambda)\widetilde{\mathbf{p}} + \frac{\lambda}{Q}\right)^{-1}, \quad \mathbb{E}[\mathbf{w}] = \sum_q \widetilde{\mathbf{p}}_q \mathbf{w}_q = 1$$

Class Probabilities to Point Estimates

对于从一个color distribution到pixel color的函数，作者团队测试了许多方法，并进行比较。

1. take the mode，虽然这会提供鲜艳明亮的颜色，但是有的时候会出现空间色彩的不一致性。比如说下图最右边的公交车上会出现红色的斑点。
2. take the mean，这种方法虽然空间色彩一致，但是会造成去饱和的现象。
3. annealed-mean，这是文中选用的方法。

下面是三种方法的一种对比。本文中采用Annealed-Mean方法，通过重新调整softmax分布中的T，然后再取平均得到最终的结果。T越低，分布就越集中。当T趋于0的时候，结果会是一个one-hot。作者发现T=0.38时能在保持色彩鲜艳度的同时，保证的空间色彩的连续性。

$$\mathcal{H}(\mathbf{Z}_{h,w}) = \mathbb{E}\big[f_T(\mathbf{Z}_{h,w})\big], \quad f_T(\mathbf{z}) = \frac{\exp(\log(\mathbf{z})/T)}{\sum_q \exp(\log(\mathbf{z}_q)/T)}$$

Fig. 4. The effect of temperature parameter T on the *annealed-mean* output (Equation 5). The left-most images show the means of the predicted color distributions and the right-most show the modes. We use $T = 0.38$ in our system.

最终的mapping函数**F**就是CNN **G**与annealed-mean **H**的复合

5、实验结果

同时为了比较不同loss function下训练的结果，作者团队用不同的loss function训练CNN，然后进行对比。

对比的指标包括：

- Perceptual realism (AMT)
- Semantic interpretability (VGG classication)
- Raw accuracy (AuC)分为class-balanced variant这一变体rebal，和non-rebal。

这些指标都是越高越好。从下表中可以看出，本文中的方法在rebal AuC和AMT指标上都达到了最好的表现。

Colorization Results on ImageNet

Method	Model Params (MB)	Model Feats (MB)	Runtime (ms)	AuC non-rebal (%)	AuC rebal (%)	VGG Top-1 Class Acc (%)	AMT Labeled Real (%)
Ground Truth	–	–	–	100	100	68.3	50
Gray	–	–	–	89.1	58.0	52.7	–
Random	–	–	–	84.2	57.3	41.0	13.0±4.4
Dahl [2]	–	–	–	90.4	58.9	48.7	18.3±2.8
Larsson et al. [23]	588	495	122.1	**91.7**	65.9	**59.4**	**27.2±2.7**
Ours (L2)	129	127	17.8	91.2	64.4	54.9	21.2±2.5
Ours (L2, ft)	129	127	17.8	91.5	66.2	56.5	23.9±2.8
Ours (class)	129	142	22.1	91.6	65.1	56.6	25.2±2.7
Ours (full)	129	142	22.1	89.5	**67.3**	56.0	**32.3±2.2**

三、代码

无