2020

Attention Branch Network: Learning of Attention Mechanism for Visual Explanation

汇报人：Lee

日期： 2020年10月18日

目录

CONTENTS

01

PART 01

第一部分

背景与动机

动机

- 使人们理解CNN的决策
  - ✓ 得到客户的认可
  - ✓ 研究模型的泛化能力、安全性问题及模型局限
  - ✓ 追溯模型预测结果
- 提高分类器的表现
  - ✓ 通过注意力机制自动去除干扰特征
  - ✓ 指导数据增强策略

ABN的主要贡献

- 提高CNN的分类效果的首次尝试
- 适用于多种基线模型，可多任务训练
- 用注意力激活图可视化

已有的一些工作

- ✓ CAM
- ✓ Grad-CAM
- ✓ Grad-CAM ++
- ✓ interpretable CNN

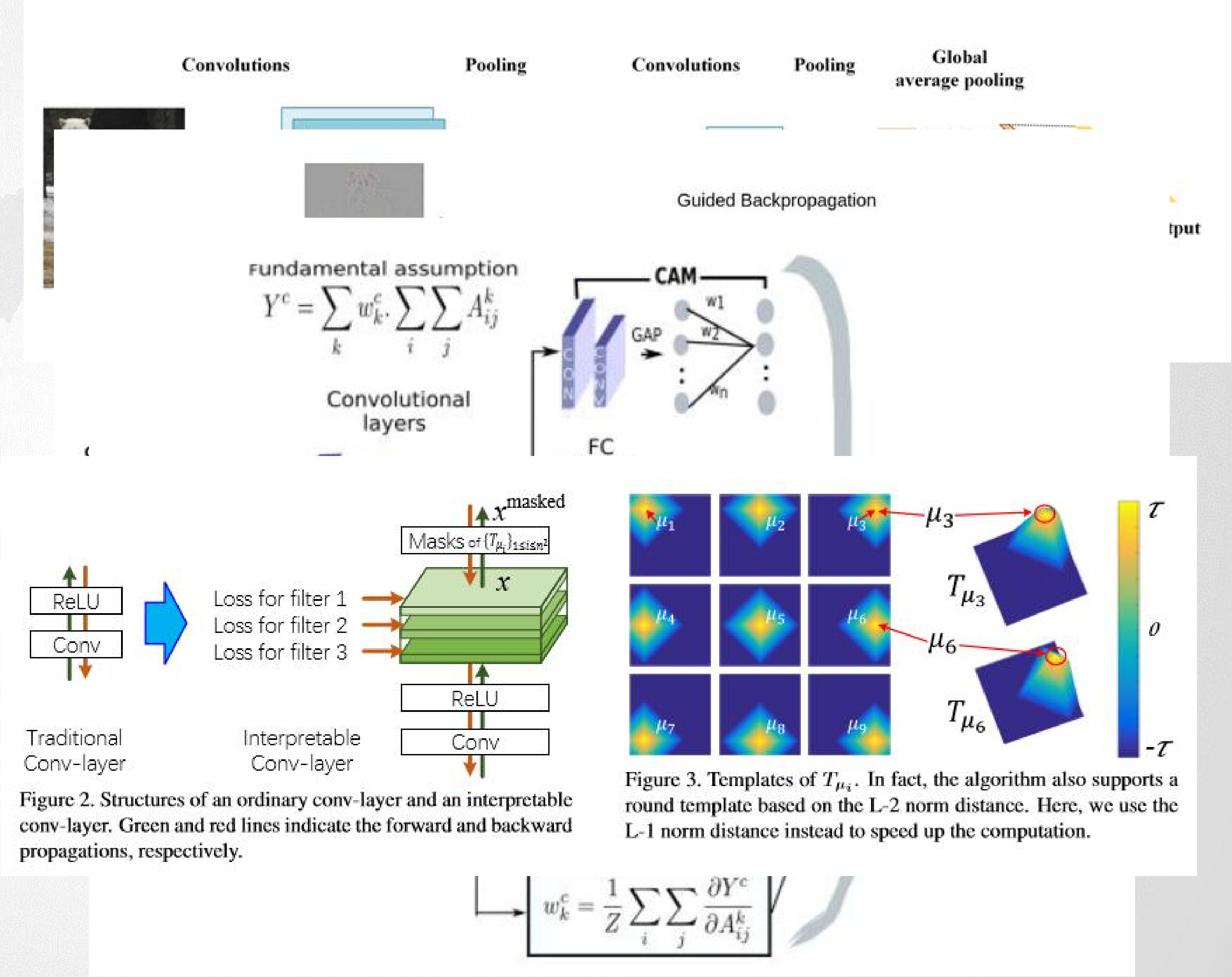


Figure 2. Structures of an ordinary conv-layer and an interpretable conv-layer. Green and red lines indicate the forward and backward propagations, respectively.

Figure 3. Templates of T_{μ_i}. In fact, the algorithm also supports a round template based on the L-2 norm distance. Here, we use the L-1 norm distance instead to speed up the computation.

ABN与CAM和Grad-CAM区别

- 不用像Grad-CAM那样要做反向传播
- 可以使用全连接网络，提高CNN的表达能力

02

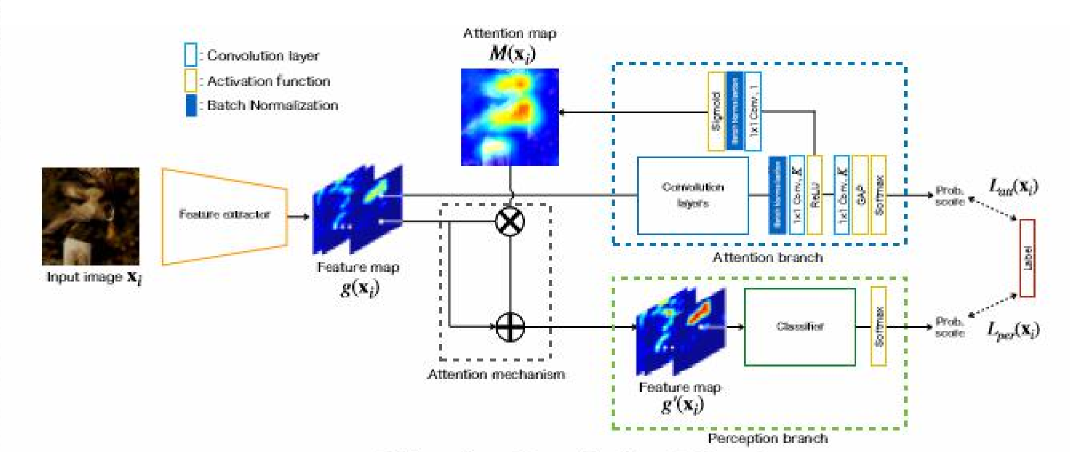
PART 02

第二部分

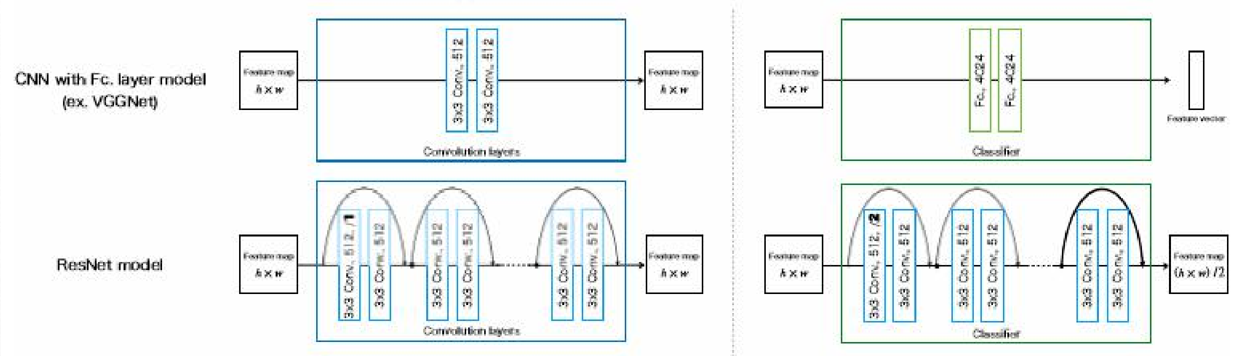
网络结构与实现原理

网络结构

- ➢ 特征提取器
- ➢ 注意力分支
- ➢ 感知分支



(a) Overview of Attention Branch Network



(b) Structure of attention branch

(c) Structure of perception branch

Figure 2. Detailed structure of Attention Branch Network.

Loss

$$L(\mathbf{x}_i) = L_{att}(\mathbf{x}_i) + L_{per}(\mathbf{x}_i)$$

$L_{att}(x_i)$ denotes training loss at the attention branch with an input sample x_i, and $L_{per}(x_i)$ denotes training loss at the perception branch.

多任务学习

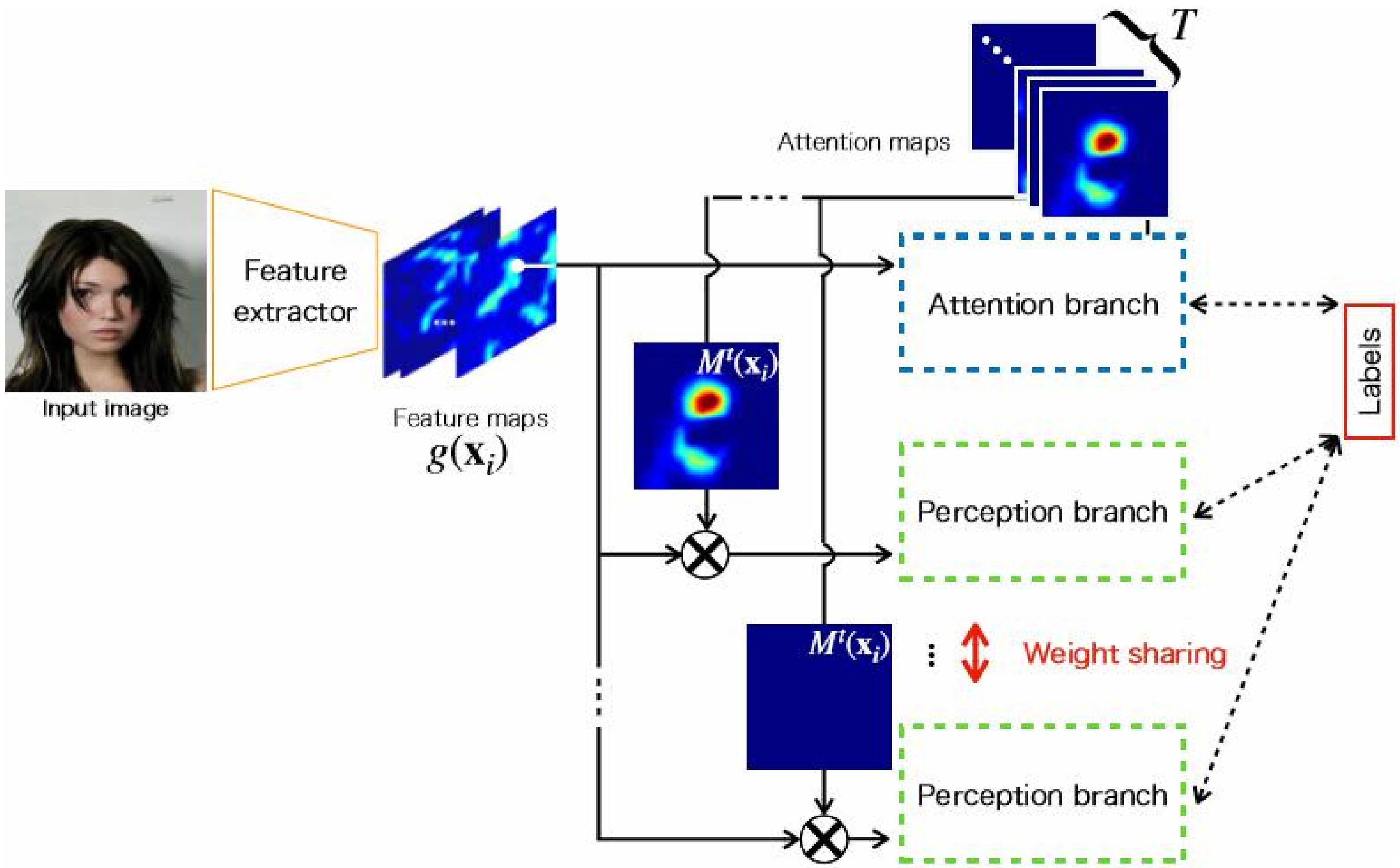


Figure 3. ABN for multi-task learning.

03

PART 03

第三部分

实验结果

图像分类

Table 1. Comparison of the top-1 errors on CIFAR100 with attention mechanism.

	$g(\mathbf{x})$	$g(\mathbf{x}) \cdot M(\mathbf{x})$	$g(\mathbf{x}) \cdot (1 + M(\mathbf{x}))$
ResNet20	31.47	30.61	**30.46**
ResNet32	30.13	28.34	**27.91**
ResNet44	25.90	**24.83**	25.59
ResNet56	25.61	24.22	**24.07**
ResNet110	24.14	23.28	**22.82**

图像分类

Table 2. Comparison of top-1 errors on CIFAR10, CIFAR100, SVHN, and ImageNet dataset.

Dataset	CIFAR10	CIFAR100	SVHN [23]	ImageNet [5]
VGGNet [14]	–	–	–	31.2
VGGNet+BN	–	–	–	26.24*
ResNet [9]	6.43	24.14*	2.18*	22.19*
VGGNet+CAM [41]	–	–	–	33.4
VGGNet+BN+CAM	–	–	–	27.42* (+1.18)
ResNet+CAM	–	–	–	22.11* (−0.08)
WideResNet [38]	4.00	19.25	2.42*	21.9
DenseNet [11]	4.51	22.27	2.07*	22.2
ResNeXt [34]	3.84*	18.32*	2.16*	22.4
Attention [32]	3.90	20.45	–	21.76
AttentionNeXt [32]	–	–	–	21.20
SENet [12]	–	–	–	21.57
VGGNet+BN+ABN	–	–	–	25.55 (−0.69)
ResNet+ABN	4.91 (−1.52)	22.82 (−1.32)	**1.86 (−0.32)**	21.37 (−0.82)
WideResNet+ABN	**3.78 (−0.22)**	18.12 (−1.13)	2.24 (−0.18)	–
DenseNet+ABN	4.17 (−0.34)	21.63 (−0.64)	2.01 (−0.06)	–
ResNeXt+ABN	3.80 (−0.04)	**17.70 (−0.62)**	2.01 (−0.15)	–
SENet+ABN	–	–	–	**20.77 (−0.80)**

∗ *indicates results of re-implementation accuracy*

图像分类

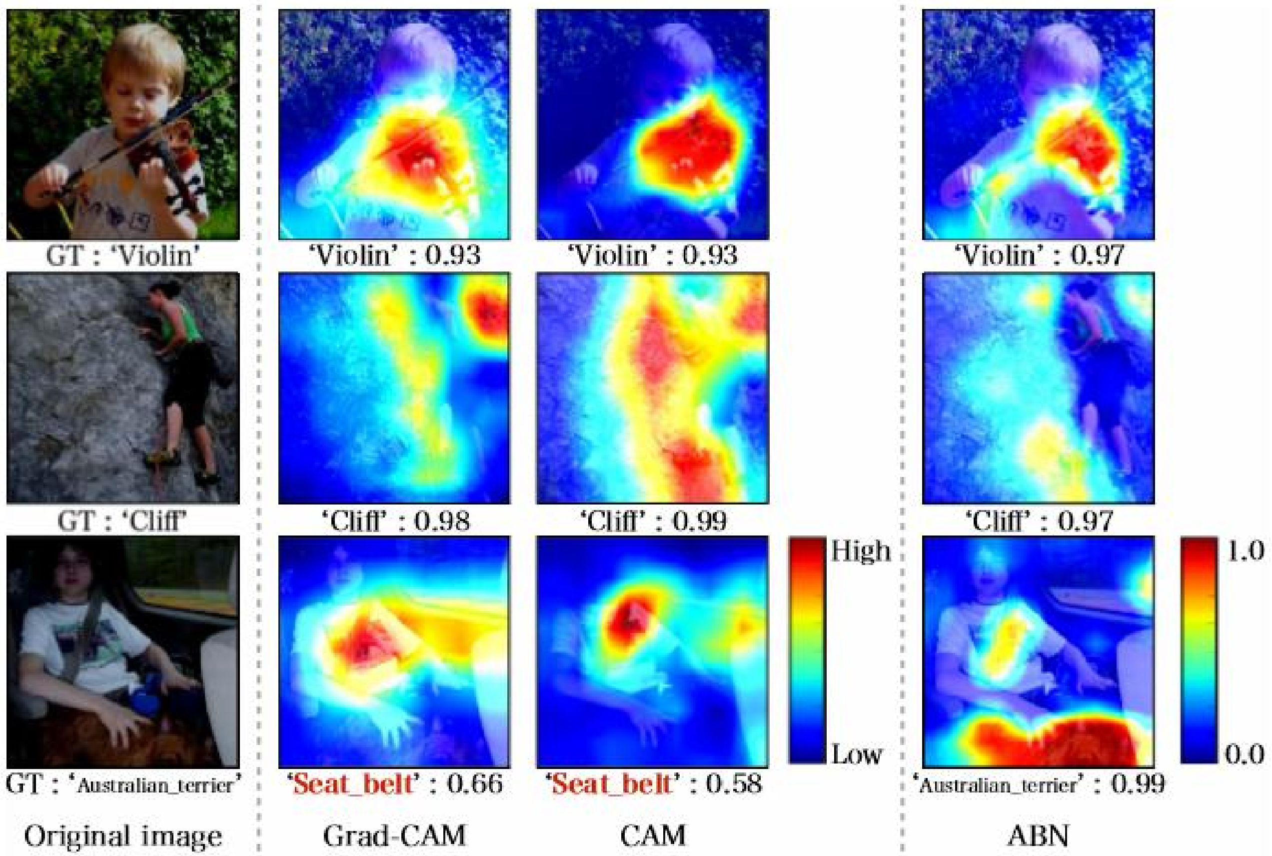


Figure 4. Visualizing high attention area with CAM, Grad-CAM, and our ABN. CAM and Grad-CAM are visualized attention maps at top-1 result.

细粒度识别

Table 3. Comparison of car model and maker accuracy on Comp-Cars dataset

task	model [%]	maker [%]
VGG16	85.9	90.4
ResNet101	90.2	90.1
VGG16+ABN	90.7	92.9
ResNet101+ABN	**97.1**	**98.1**

细粒度识别

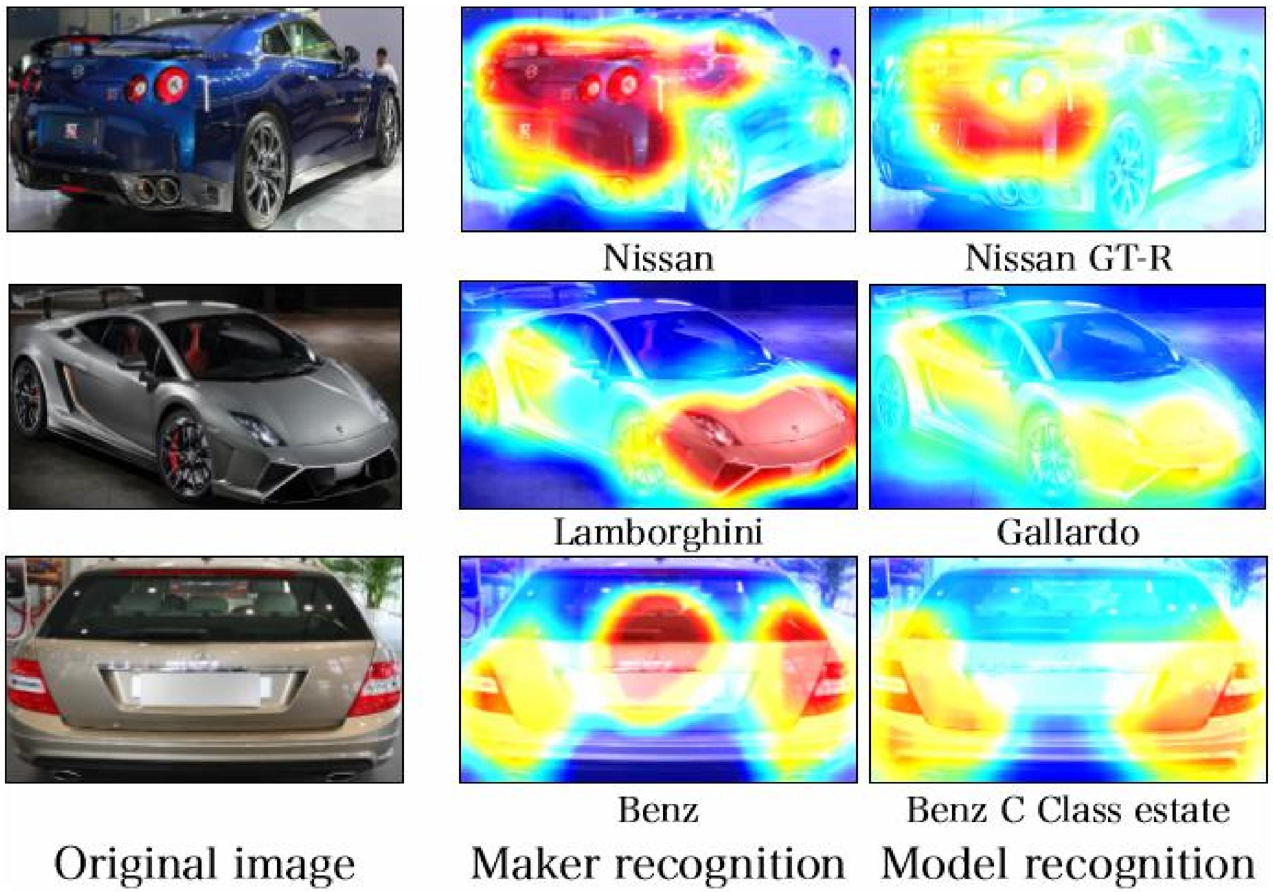


Figure 5. Visualizing attention map on fine-grained recognition.

多任务学习

Table 4. Comparison of multiple facial attribute recognition accuracy on CelebA dataset

Method	Average of accuracy [%]	Odds
FaceTracer [16]	81.13	40/40
PANDA-1 [40]	85.43	39/40
LNet+ANet [42]	87.30	37/40
MOON [28]	90.93	29/40
ResNet101	90.69	27/40
ABN	**91.07**	–

多任务学习

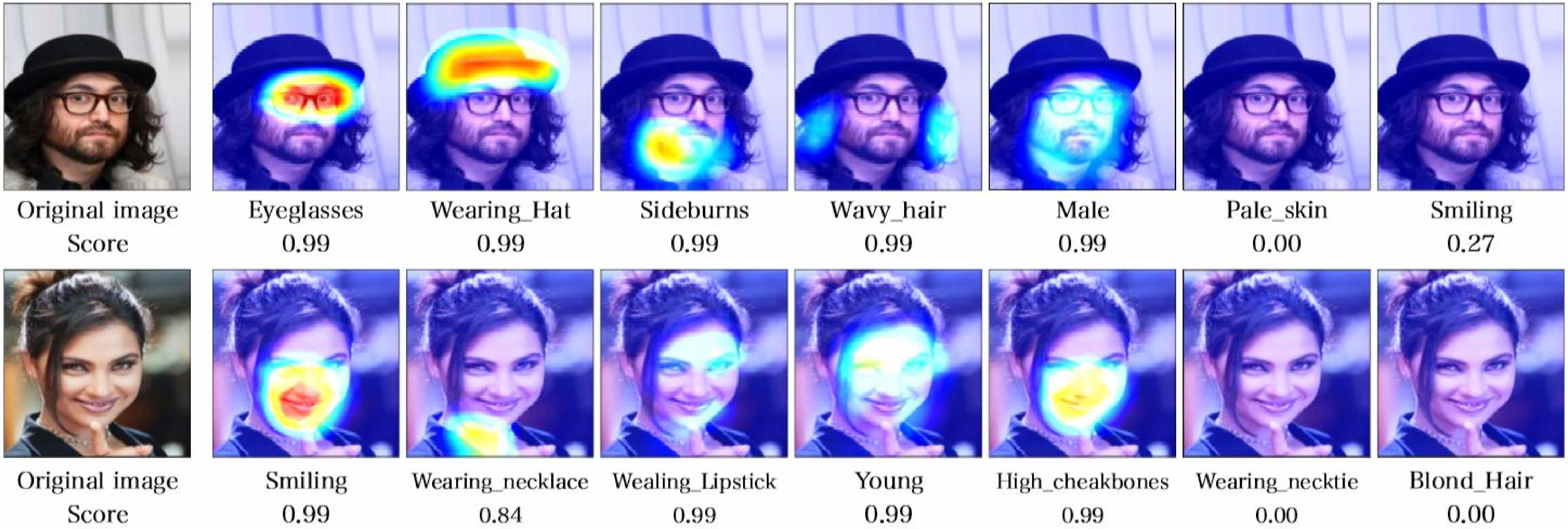


Figure 7. Visualizing attention maps on multiple facial attributes recognition. These scores are final recognition scores at the perception branch.

04

PART 04

第四部分

问题与讨论

谢谢！