

法律声明

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象

■ 新浪微博：ChinaHadoop



卷积神经网络—目标分类

主讲人： 李伟

纽约城市大学博士

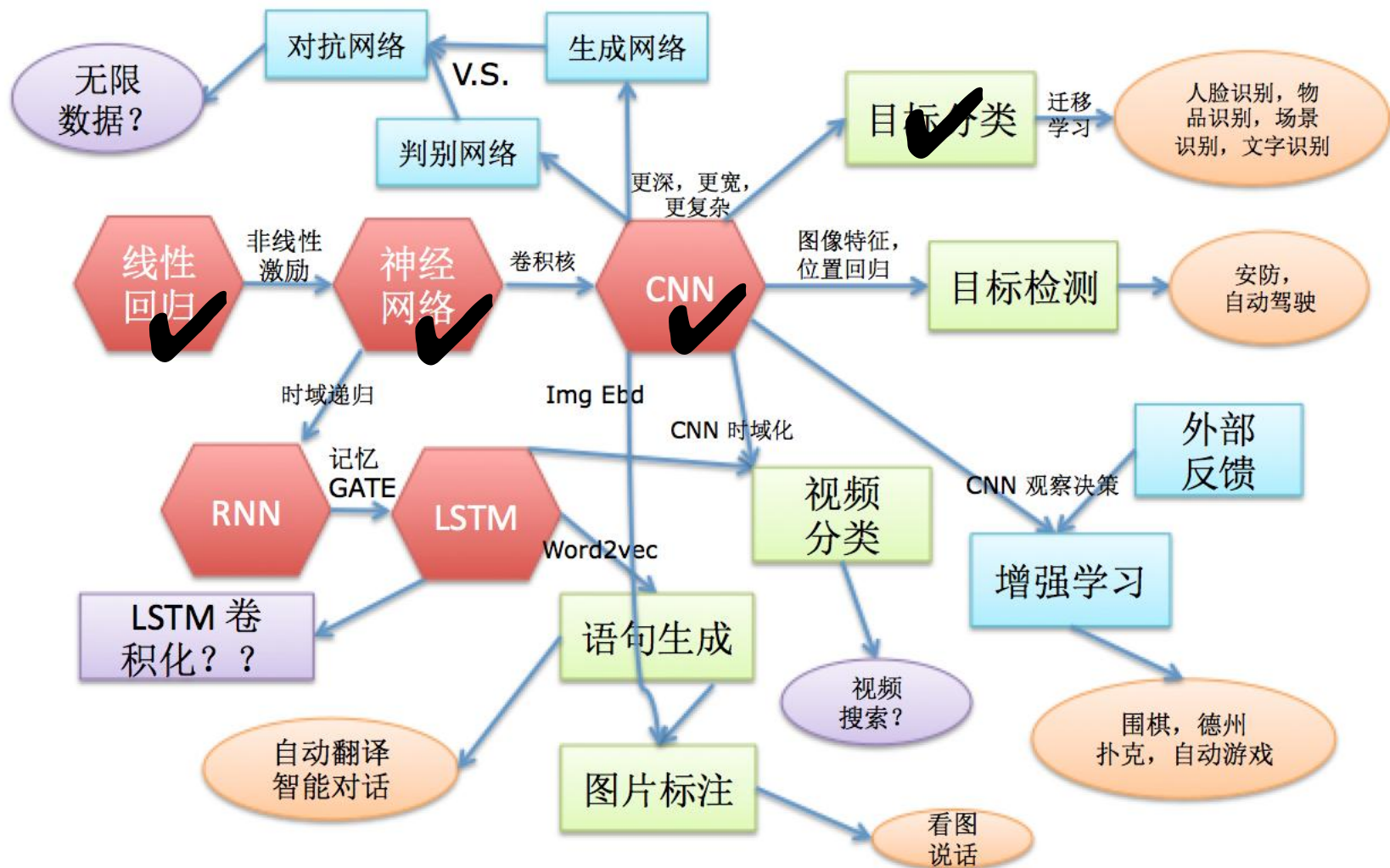
主要研究深度学习，计算机视觉，人脸计算
多篇重要研究文章作者，重要会议期刊审稿人

微博ID: weightlee03（相关资料分享）

GitHub ID: wiibrew（课程代码发布）

<https://github.com/wiibrew/DeepLearningCourseCodes>

结构



提纲

- 目标分类基本框架
- 迁移学习
- 研究分享：如何设计神经网络
- 实例：基于VGG进行人脸表情识别

期待目标

- 1. 知道目标分类的实现过程
- 2. 明白迁移学习在目标分类过程中的作用
- 3. 了解如何设计神经网络结构针对特殊结构的目标进行分类
- 4. 能够完成表情识别实例训练：数据准备，模型设计／更改，参数调整，数据测试

提纲

- 目标分类基本框架
- 迁移学习
- 研究分享：如何设计神经网络
- 实例：基于VGG进行人脸表情识别

目标分类基本框架

□ 1. 数据准备

□ 2. 模型设计

□ 3. 训练细节

目标分类基本框架

□ 1. 数据准备

数据来源

现有数据集的子集；网络采集；现有数据人工标注

数据扩充

原始数据切割；噪声颜色等像素变化；旋转平移等姿态变化

数据规范

均值处理；归一化；大小调整

目标分类基本框架

□ 1. 数据准备

数据来源

现有数据: <http://deeplearning.net/datasets/>

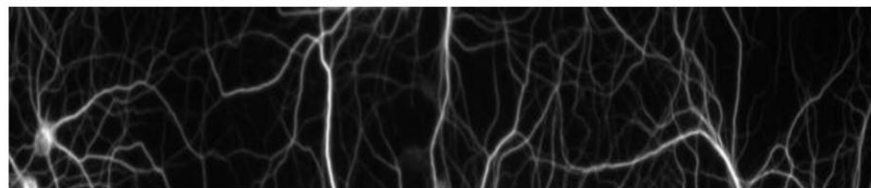
自然图片

人工合成图片

人脸

文本

对话



Recent Posts

MILA is Hiring Two Software Engineers
OpenAI: A new non-profit AI company

Datasets

These datasets can be used for benchmarking deep learning algorithms:

Symbolic Music Datasets

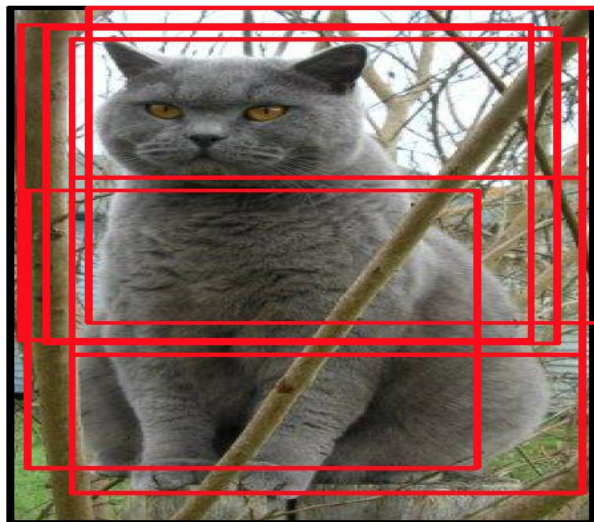
目标分类基本框架

□ 1. 数据准备

数据扩充

原始数据切割；噪声颜色等像素变化；旋转平移等姿态变化

局部切割：

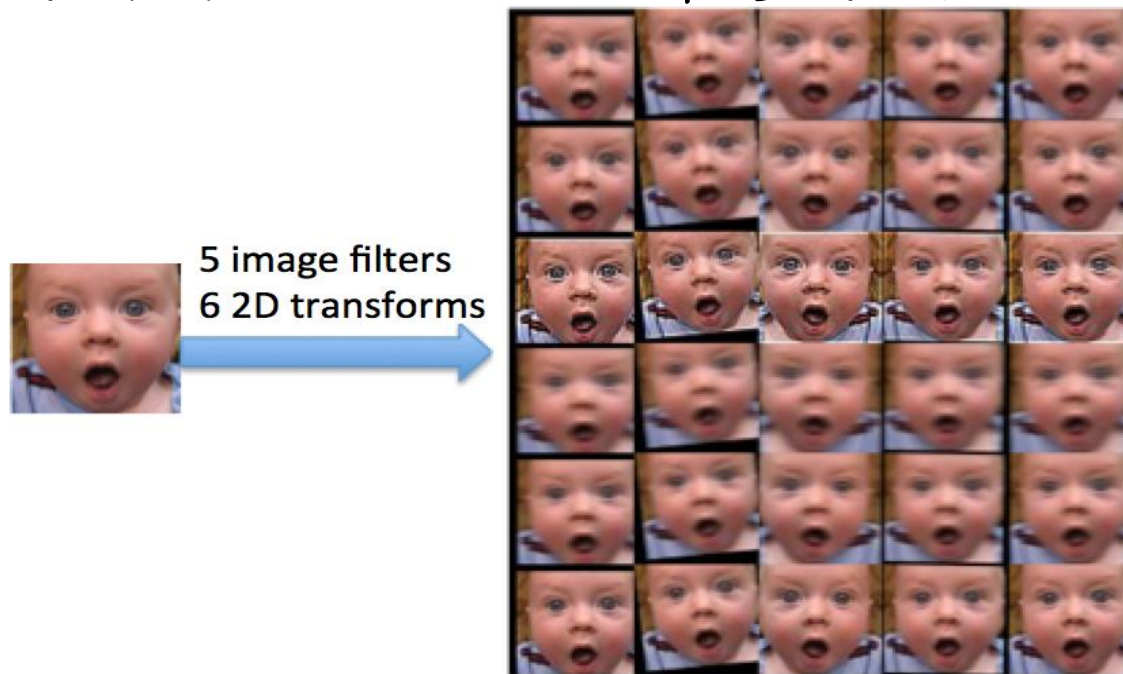


目标分类基本框架

□ 1. 数据准备

数据扩充

噪声 颜色等像素变化； 旋转 平移等姿态变化



目标分类基本框架

□ 数据准备

□ 图像效果算子

更模糊，更清楚(unsharp)，动作模糊(motion)

□ 旋转平移R, T

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos a & -\sin a & h \\ \sin a & \cos a & k \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos a & \sin a & -h \cos a - k \sin a \\ -\sin a & \cos a & h \sin a - k \cos a \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}$$

目标分类基本框架

□ 1. 数据准备

数据扩充(matlab)

效果卷积核

旋转矩阵

```
F1 = fspecial('disk',1);
F2 = fspecial('average',4);
F3 = fspecial('gaussian',3);
F4 = fspecial('gaussian',4);
F5 = fspecial('unsharp');
F6 = fspecial('motion',7,0.75*pi);
F7 = fspecial('motion',5,0);
F8 = fspecial('motion',5,pi/2);
F9 = fspecial('motion',6,pi);
F10 = [0 0 0; 0 1 0; 0 0 0];
vfilter = {F1,F3, F5, F6, F9,F10};
```

```
vTrans ={};
t1=maketform('affine',[1 0 2*f; 0 1 2*f; 0 0 1]);
t2=maketform('affine',[0.96 0.1 0; -0.1 0.96 0; 0 0 1]);
t3=maketform('affine',[-1 0 0; 0 1 0; w1 0 1]);
t4=maketform('affine',[0.97 -0.02 0; 0.02 0.97 0; 0 0 1]);
t5=maketform('affine',[1 0 0; 0 1 0; 0 0 1]);
t6=maketform('affine',[1.04 -0.1 0; 0.1 1.04 0; 0 0 1]);
t7=maketform('affine',[1.03 0.02 0; -0.02 1.03 0; 0 0 1]);

vTrans = {t1, t2, t3, t5, t6};
```

目标分类基本框架

□ 1. 数据准备

数据规范

均值处理；归一化；大小调整

```
# ts_data=ts_data[:,:,:,:16:240]
|tr_data=tr_data-MEAN_IMAGE
# ts_data=ts_data-MEAN_IMAGE
```

```
# trdata=trdata
trdata/=256
return train_fr
```

```
for t in range(3):
    im224[t,:,:]=cv2.resize(im1[:,:,:t],(224,224))
datablob[i,:,:,:]=im224
```

目标分类基本框架

□ 2. 模型设计

任务类型

现有模型(The-state-of-the-art)

局部更改-从头设计

目标分类基本框架

□ 2. 模型设计

任务类型

分类： 表情分类，属于什么种类，人群分类

分类 + 回归： 表情 + 程度，种类 + 信心，什么人 + 人数

多目标分类： 面部行为，群体行为，车流预测

目标分类基本框架

□ 2. 模型设计

□ 现有模型(The-state-of-the-art)

已有的方案能否／怎样借鉴

ICCV, ECCV, CVPR

ICML NIPS,

ACL, KDD



Computer Science

- [Computing Research Repository \(CoRR new, recent, find\)](#)

includes (see [detailed description](#)): Artificial Intelligence; Computation and Language; Computational Complexity; Computational Engineering, Finance, and Science; Computational Geometry; Computer Science and Game Theory; Computer Vision and Pattern Recognition; Computers and Society; Cryptography and Security; Data Structures and Algorithms; Databases; Digital Libraries; Discrete Mathematics; Distributed, Parallel, and Cluster Computing; Emerging Technologies; Formal Languages and Automata Theory; General Literature; Graphics; Hardware Architecture; Human-Computer Interaction; Information Retrieval; Information Theory; Learning; Logic in Computer Science; Mathematical Software; Multiagent Systems; Multimedia; Networking and Internet Architecture; Neural and Evolutionary Computing; Numerical Analysis; Operating Systems; Other Computer Science; Performance; Programming Languages; Robotics; Social and Information Networks; Software Engineering; Sound; Symbolic Computation; Systems and Control

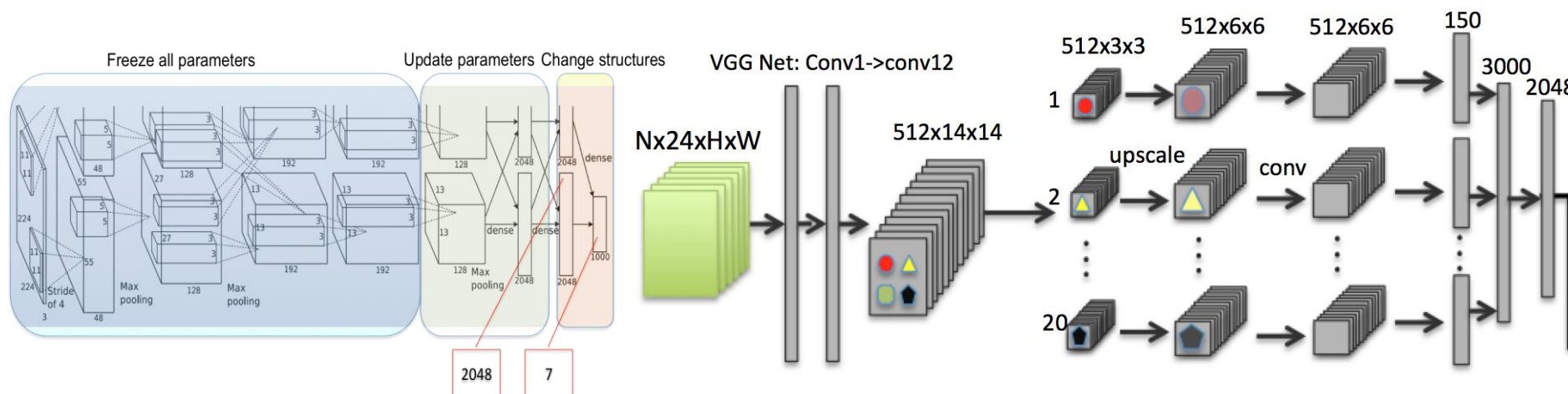
目标分类基本框架

□ 2. 模型设计

局部更改-从头设计

更改：哪里改变；新参数确定

设计：新结构特点，为什么可行



目标分类基本框架

□ 3. 训练细节

- GPU-Batch size, 是否并行
- 数据循环方式 / 平衡性考虑
- 网络深度宽度确定
- 损失函数设计
- 学习率变化方式, 模型各层学习率是否一致
- 评价方式: 准确率, F1 score

目标分类基本框架

□ 3. 训练细节

□ GPU-Batch size, 是否并行
GPU内存 - Batch Size关系

□ 数据循环方式 / 平衡性考虑

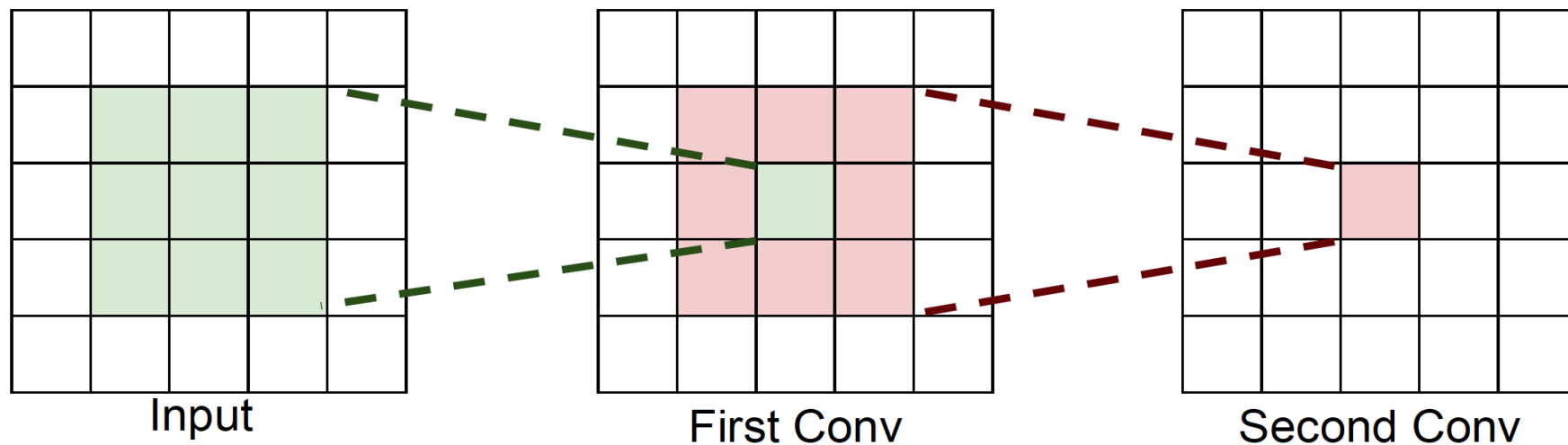
1. 数量较少的类别, 数据是否需要补偿
2. 从头到尾多次循环
2. 每次随机选取部分数据

目标分类基本框架

□ 3. 训练细节

网络深度宽度确定

深度与卷积核关系：第 i 层卷积核关注范围？



层数变多，参数变少。

目标分类基本框架

□ 3. 训练细节

□ 损失函数设计

分类：SOFTMAX， 直接拟合

□ 学习率变化方式， 模型各层学习率是否一致

目标分类基本框架

□ 3. 训练细节

□ 评价方式：准确率，F1 score

精度： 准确估计 / 总数

0/1 分类中

F1 score: $2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$

Recall: 正确的1识别 / 真值所有1个数

Precision: 正确的1的识别 / 所有认为是1的个数

目标分类基本框架

□ 1. 数据准备

□ 2. 模型设计

□ 3. 训练细节

提纲

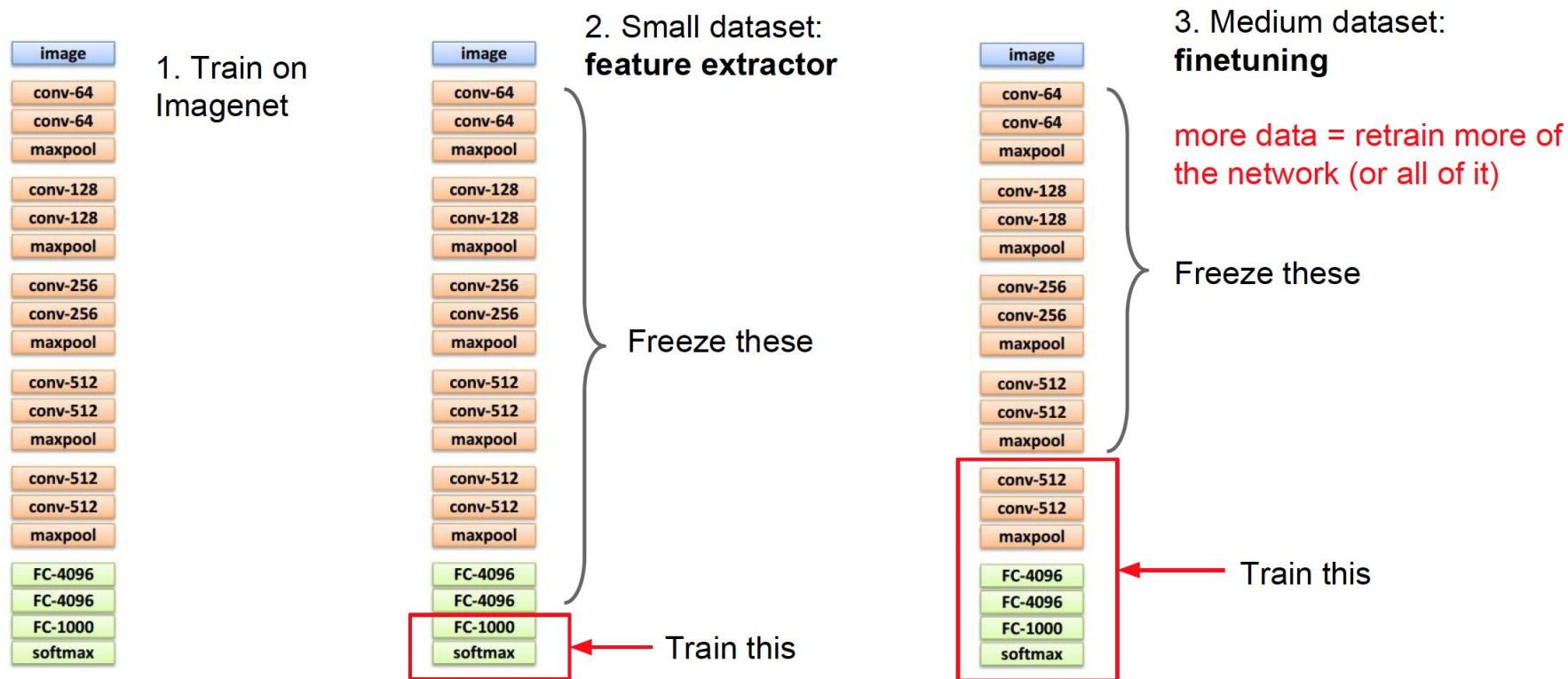
- 目标分类基本框架
- 迁移学习
- 研究分享：如何设计神经网络
- 实例：基于VGG进行人脸表情识别

迁移学习

□ 问题：ImageNet 上亿参数，数据量百万，是不是参数多的模型都需要大量数据？

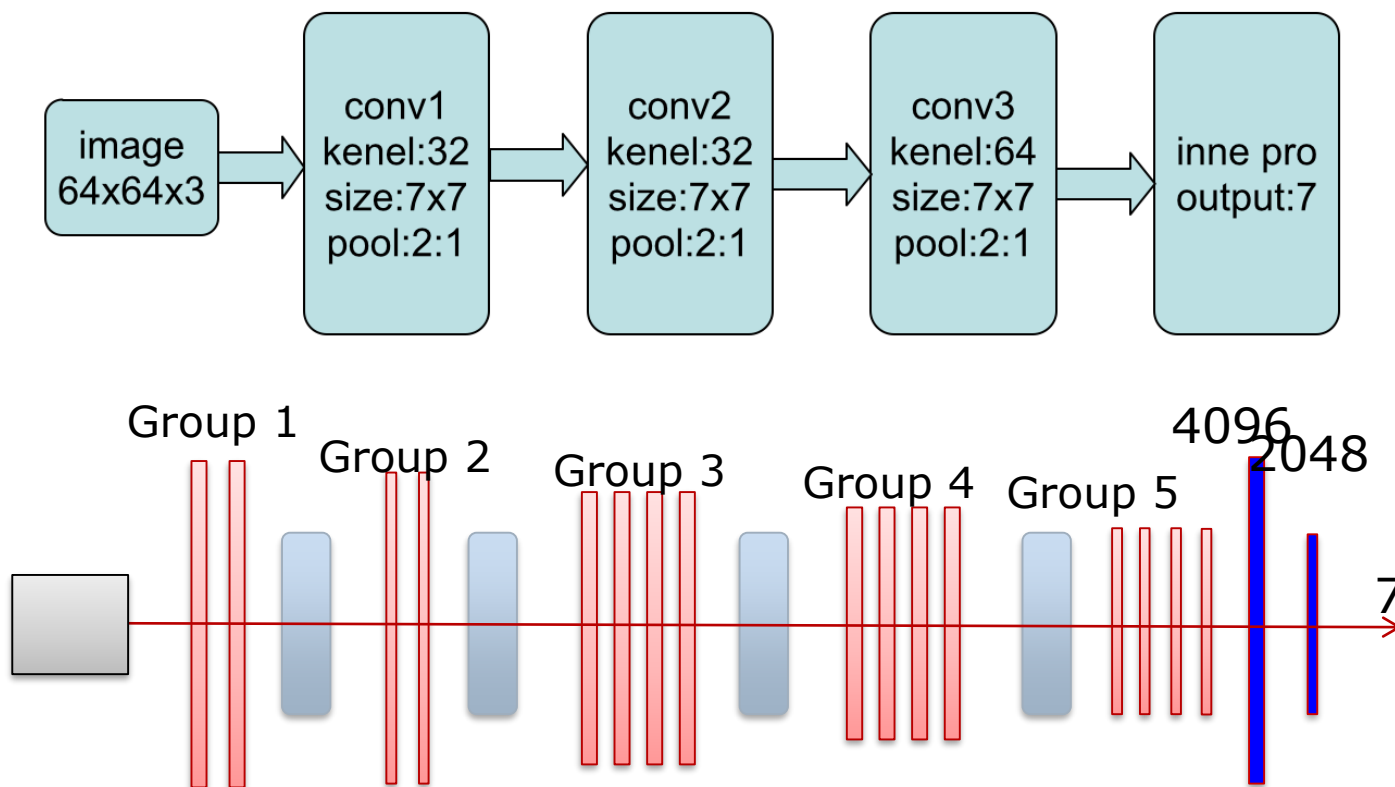
迁移学习

□ 问题：ImageNet 上亿参数，数据量百万，是不是参数多的模型都需要大量数据？



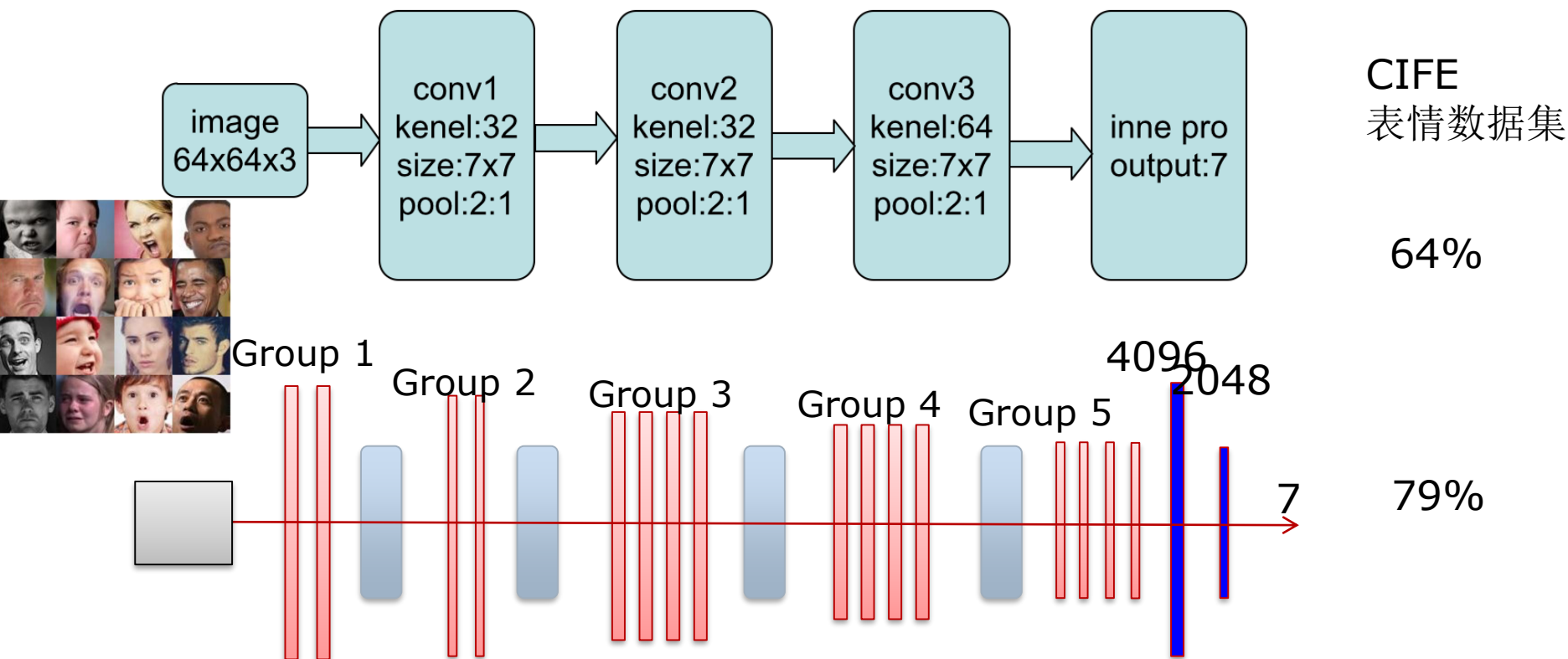
迁移学习

□ 实验对比 - 是否有用?



迁移学习

□ 实验对比 - 是否有用?



迁移学习

□ 不同数据处理

| | 类似数据 | 不同数据 |
|------|-------|-------|
| 少量数据 | FC层更新 | ??? |
| 大量数据 | 中高层更新 | 更新更多层 |

□ 学习率的处理

□ 最低卷积层基本不变

□ 中间卷积层看情况

□ 最后全连接，结构参数均变化

迁移学习

□ 基础模型选择

是否已有特定任务的模型

ImageNet 模型：VGG，GoogLeNet，ResNet

只是简单分类？

大量结构设计？

提纲

- 目标分类基本框架
- 迁移学习
- 研究分享：如何设计神经网络
- 实例：基于VGG进行人脸表情识别

研究分享： 如何设计神经网络

□ 研究问题： 如何进行面部行为识别 (AU detection)

面部行为单元：



eyebrows lower
tighter [1]

cheek raiser

chin raiser

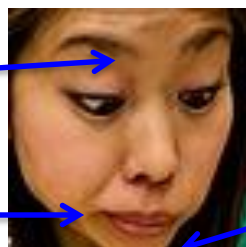
lip

Inner Brow Raiser

Outer Brow Raiser

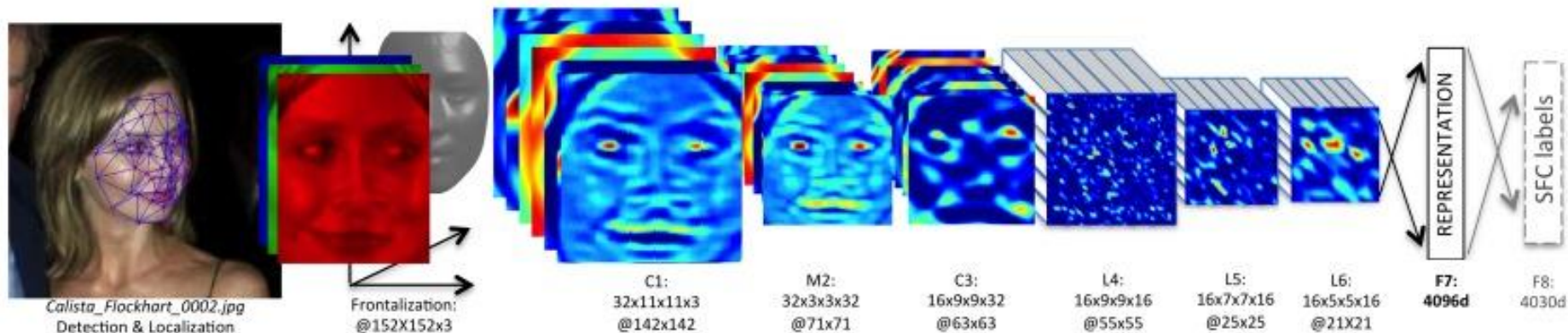
Lip Pressor

**Lip
Tightener**



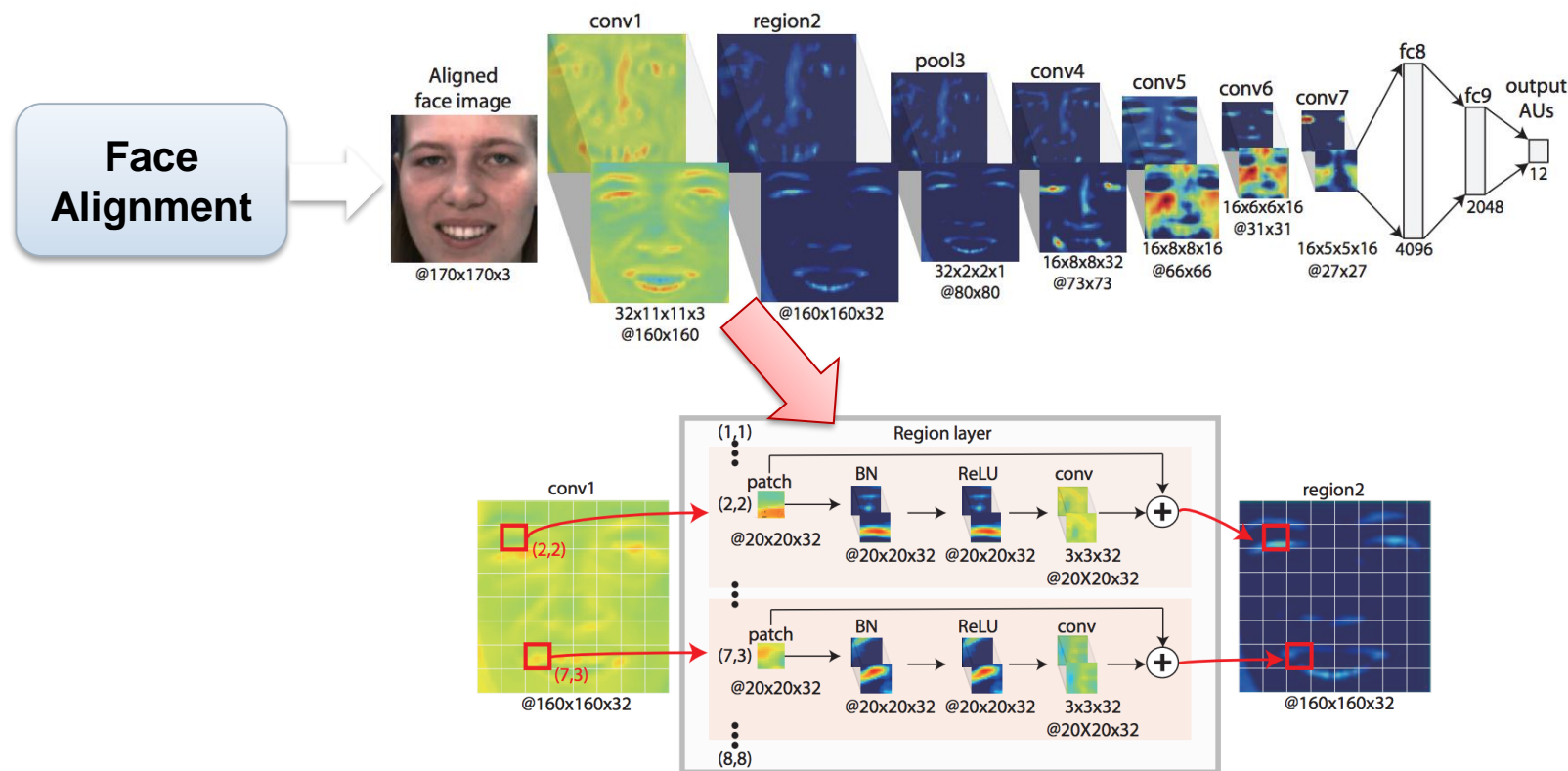
研究分享：如何设计神经网络

□ 已有方法



研究分享：如何设计神经网络

□ 已有方法



研究分享： 如何设计神经网络

□ 不足

1. 预处理：大量对准，对对准要求高，原始信息可能丢失
2. 卷积参数数量很大，模型收敛难度大，需要大量数据
3. 模型可扩展性差，基本限于人脸计算

研究分享： 如何设计神经网络

- 改进想法 - 神经网络
- 不需要预处理，自动进行局部探测
- 不要所有区域都处理，更多关注在有意义的区域
- 重要区域之间不会影响削弱学习效果

研究分享： 如何设计神经网络

- 设计来源1:
- 注意力网络 – attention layer



Original Image

First Attention Layer

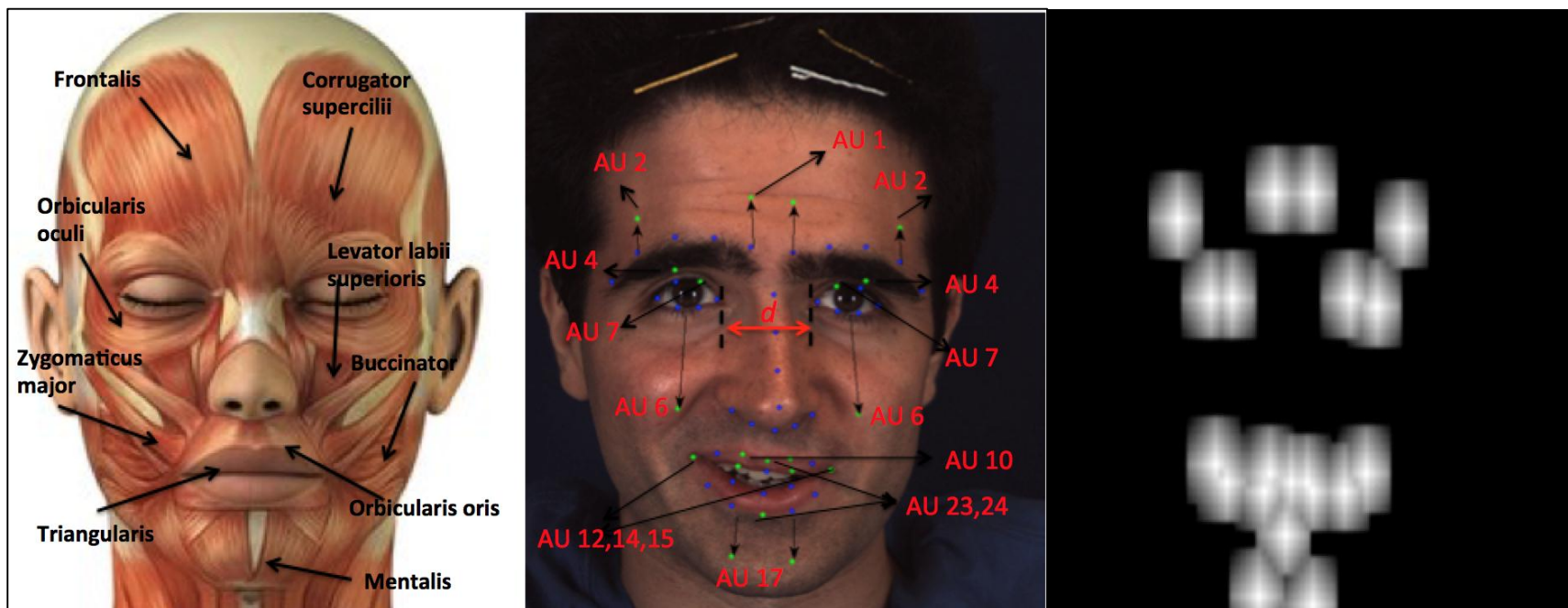
Second Attention Layer

研究分享：如何设计神经网络

□ 注意力网络

形成注意力网络

Facial muscles Subject facial landmarks Attention map



研究分享： 如何设计神经网络

□ 注意力网络

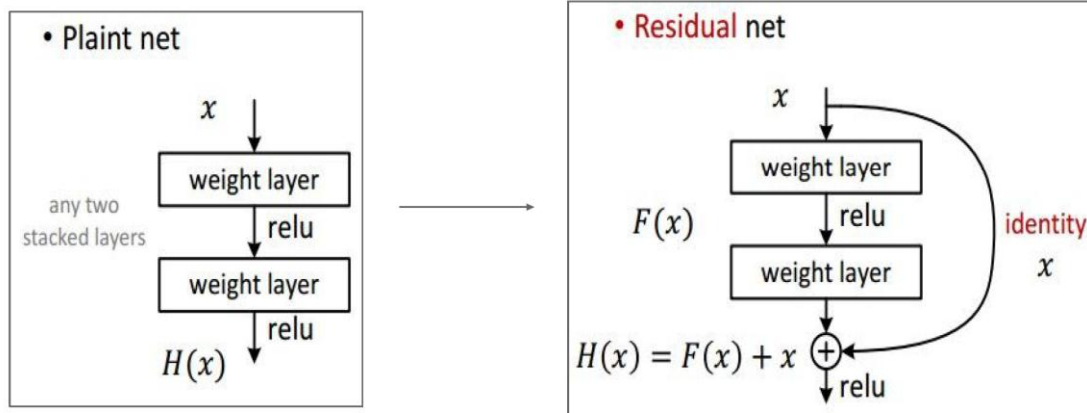
1. Dlib(或原始数据集) 找到人脸关键点
2. 人脸关键点 → 行为单元中心
3. 由中心生成注意力图

研究分享： 如何设计神经网络

□ 注意力网络

注意力网络添加：

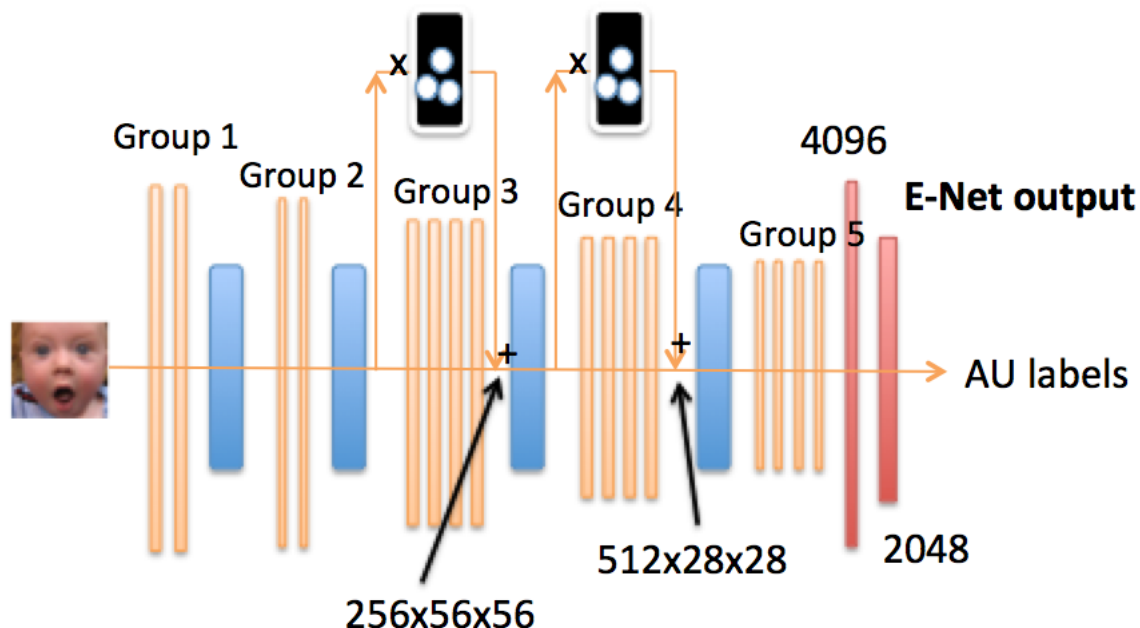
添加在哪里？ 什么方式添加？



研究分享：如何设计神经网络

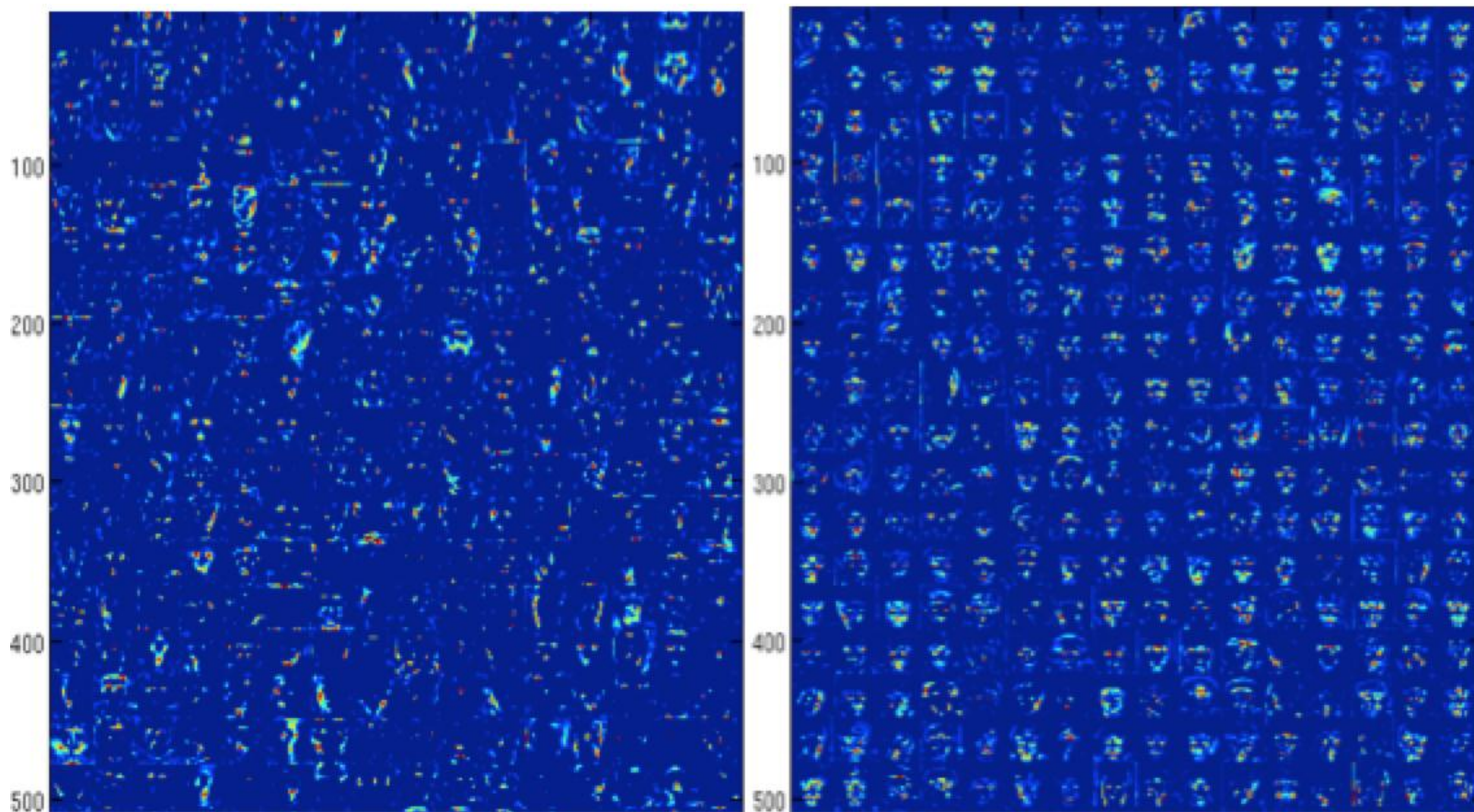
□ 注意力网络

注意力网络添加：



研究分享： 如何设计神经网络

□ 注意力网络-效果

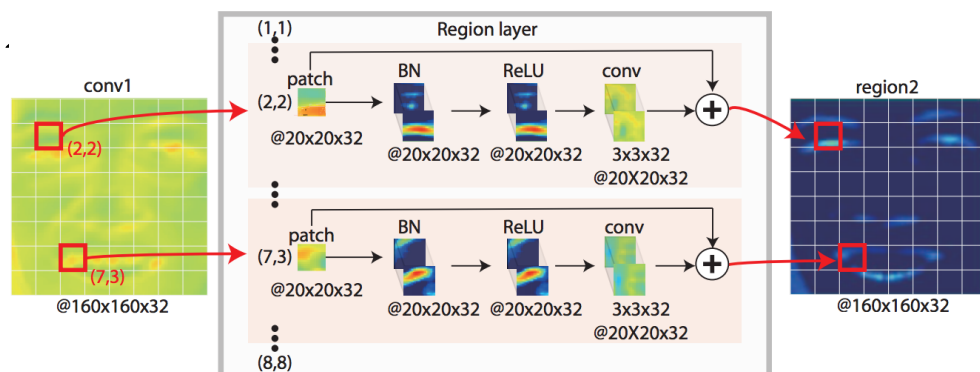


研究分享： 如何设计神经网络

□ 设计来源2:

□ 局部学习网络

针对不同的区域进行针对性学习，不同的区域的学习不

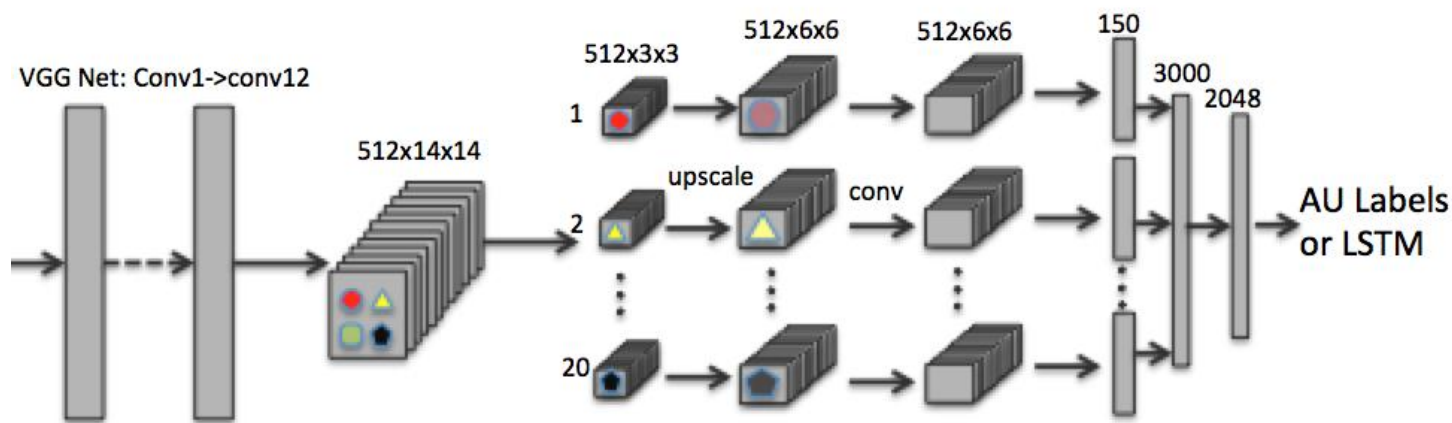


对区域的分布能够自动适应

研究分享： 如何设计神经网络

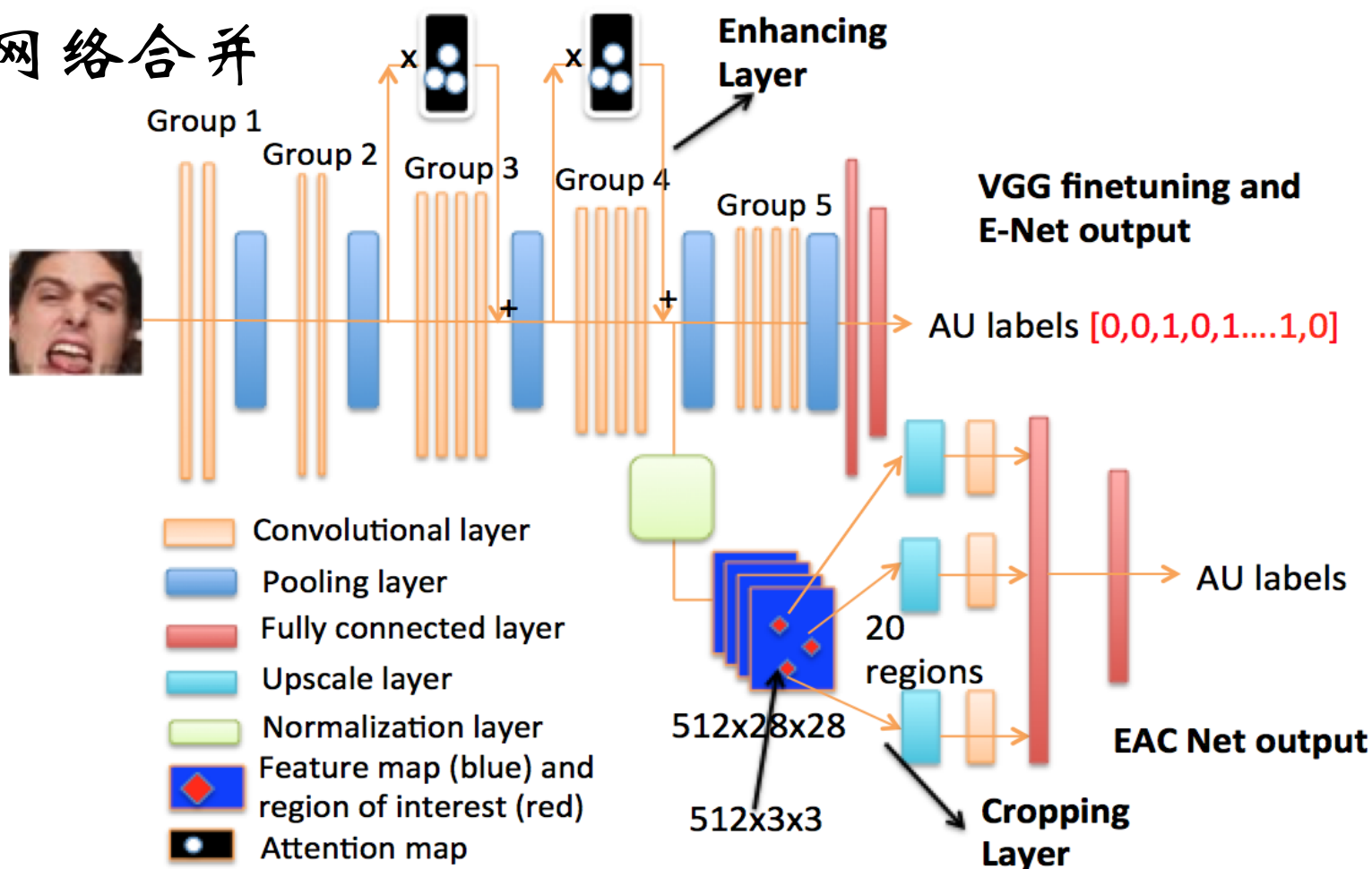
□ 局部学习网络

对区域的分布能够自动适应



研究分享：如何设计神经网络

网络合并



研究分享： 如何设计神经网络

□ 效果

| AU | LSVM | JPML[13] | DRML[21] | FVGG | E-Net | EAC |
|-----|------|-------------|-------------|------|-------------|-------------|
| 1 | 23.2 | 32.6 | 36.4 | 27.8 | 37.6 | 39.0 |
| 2 | 22.8 | 25.6 | 41.8 | 27.6 | 32.1 | 35.2 |
| 4 | 23.1 | 37.4 | 43.0 | 18.3 | 44.2 | 48.6 |
| 6 | 27.2 | 42.3 | 55.0 | 69.7 | 75.6 | 76.1 |
| 7 | 47.1 | 50.5 | 67.0 | 69.1 | 74.5 | 72.9 |
| 10 | 77.2 | 72.2 | 66.3 | 78.1 | 80.8 | 81.9 |
| 12 | 63.7 | 74.1 | 65.8 | 63.2 | 85.1 | 86.2 |
| 14 | 64.3 | 65.7 | 54.1 | 36.4 | 56.8 | 58.8 |
| 15 | 18.4 | 38.1 | 33.2 | 26.1 | 31.6 | 37.5 |
| 17 | 33.0 | 40.0 | 48.0 | 50.7 | 55.6 | 59.1 |
| 23 | 19.4 | 30.4 | 31.7 | 22.8 | 21.9 | 35.9 |
| 24 | 20.7 | 42.3 | 30.0 | 35.9 | 29.1 | 35.8 |
| Avg | 35.3 | 45.9 | 48.3 | 43.8 | 52.1 | 55.9 |

EAC-Net: A Region-based Deep Enhancing and Cropping Approach for Facial Action Unit Detection [<https://arxiv.org/abs/1702.02925>]

研究分享： 如何设计神经网络

□ 作用

- 无需提前进行面部对准就可以对面部行为识别
- 脸部各个行为单元局部针对学习，局部信息可以单独用于某个行为单元识别
- 根据控制肌肉的分布以及人脸特征点检测结果确定区域，更具有合理性以及可操作性

研究分享：如何设计神经网络

□ 总结

如何设计神经网络

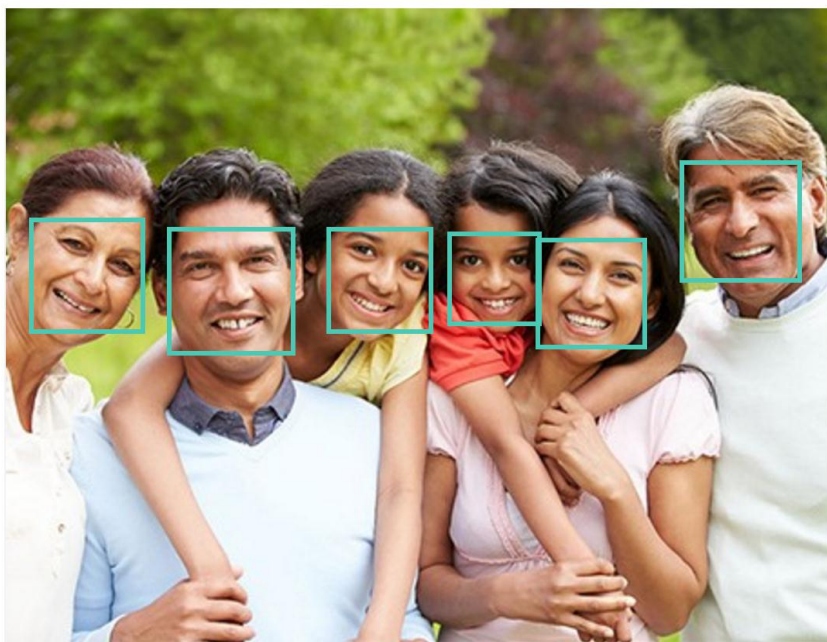
1. 明确研究问题
2. 已有的解决方案特点，借鉴地方，不足
3. 突破点在哪里，还是重复就足够了
4. 想法在网络结构中实现
5. 训练结果反馈调整

提纲

- 目标分类基本框架
- 迁移学习
- 研究分享：如何设计神经网络
- 实例：基于VGG进行人脸表情识别

实例：基于VGG进行人脸表情识别

- 表情识别
- 通过人脸分析，找出属于7类表情中哪一类，并提供分数



Detection Result:
6 faces detected

JSON:

```
[
  {
    "faceRectangle": {
      "left": 89,
      "top": 120,
      "width": 71,
      "height": 71
    },
    "scores": {
      "anger": 3.44697156e-7,
      "contempt": 6.8442256e-7,
      "disgust": 0.00009313403,
      "fear": 2.90478118e-13,
      "happiness": 0.999748647,
      "neutral": 0.000157187445,
      "sadness": 4.30214253e-10,
      "surprise": 1.04141549e-8
    }
  }
]
```

实例：基于VGG进行人脸表情识别

□ 发展现状



:) **Affectiva**

 **IMOTIONS**[®]
BIOMETRIC RESEARCH PLATFORM

 **Microsoft**

EMOTIENT

Overview

Status

Acquired by Apple on January 7, 2016

Cognitive Services

实例：基于VGG进行人脸表情识别

□ 数据集：CIFE: Candid image for facial expression
关键词搜索



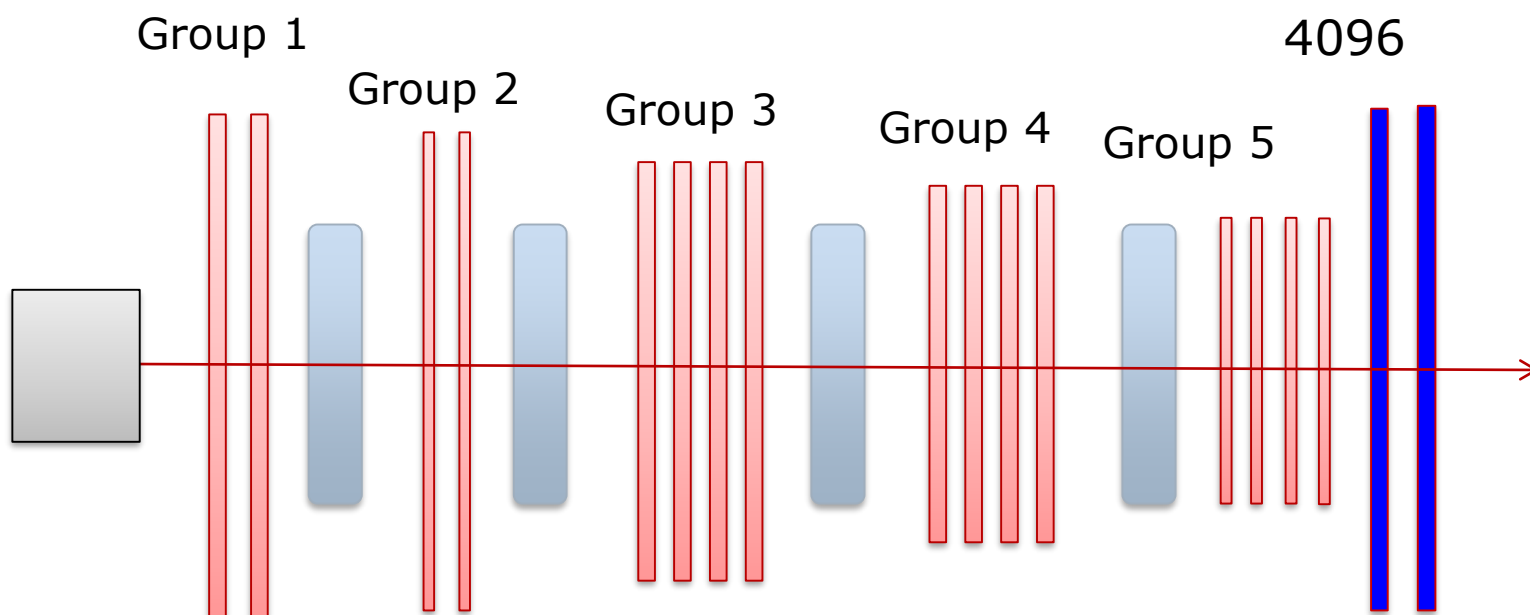
人脸探测，**label**确定



A Deep-Learning Approach to Facial Expression Recognition
with Candid Images. W Li, M Li, Z Su, Z Zhu
<https://drive.google.com/open?id=0B3ANX1iL124qbmXOc2cyQzhvUFE>

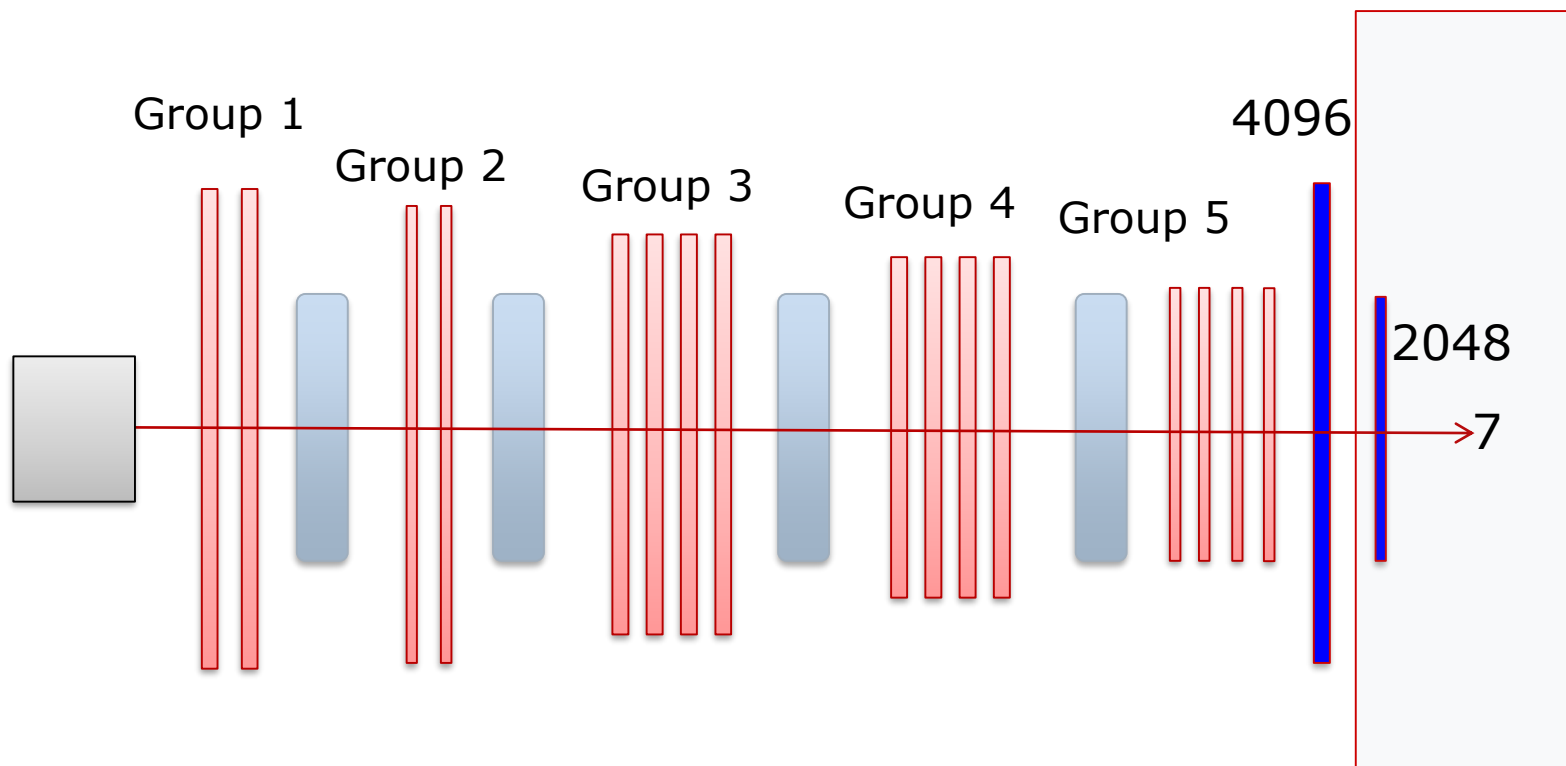
实例：基于VGG进行人脸表情识别

□ VGG模型 fine-tuning



实例：基于VGG进行人脸表情识别

□ VGG模型结构调整



总结

□ 有问题请到课后交流区

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他人回复问题

□ 讲师微博：weightlee03，每周不定期分享DL资料

□ GitHub ID：wiibrew（课程代码发布）

<https://github.com/wiibrew/DeepLearningCourseCodes>

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：小象
- 新浪微博：ChinaHadoop

