

硕士毕业论文答辩

基于深度学习的轻量化斜眼检测系统 RF-CNN Lite

答辩人：黄龙涛

指导老师：范衡

2020.06.14

- 一、背景与意义
- 二、研究状况
- 三、工作介绍
- 四、实验结果
- 五、创新与不足
- 六、总结与展望

斜视是双眼注视物体时无法正确对准的一种眼科疾病。[1]



[1] "Visual Processing: Strabismus". National Eye Institute. National Institutes of Health. June 16, 2010. Archived from the original on October 5, 2016. Retrieved October 2, 2016.

斜视会给患者带来以下危害：

- 复视 2 ；
- 弱视 3 ；
- 致盲；
- 影响外貌美观。

因此，及时对患者进行斜视检测有助于其预后以及治疗。

[2] Kiorpes, L., Kiper, D. C., O' keefe, L. P., Cavanaugh, J. R. & Movshon, J. A. Neuronal correlates of amblyopia in the visual cortex of macaque monkeys with experimental strabismus and anisometropia. J. Neurosci. 18, 6411–6424 (1998)

[3] Van de Graaf, E. et al. Amblyopia and strabismus questionnaire: design and initial validation. Strabismus 12, 181–193 (2004).

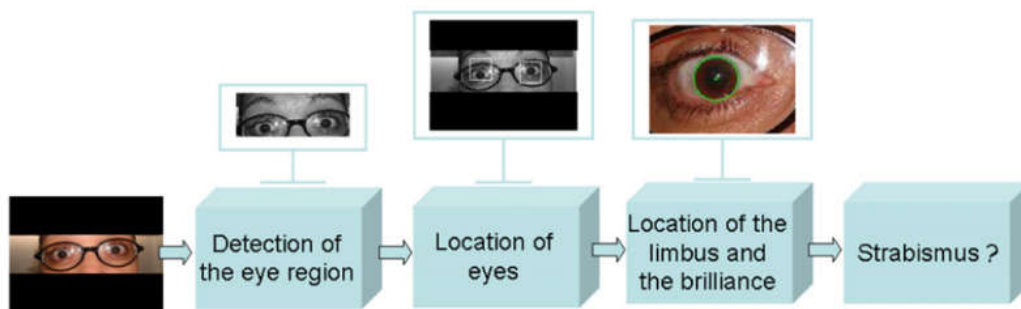
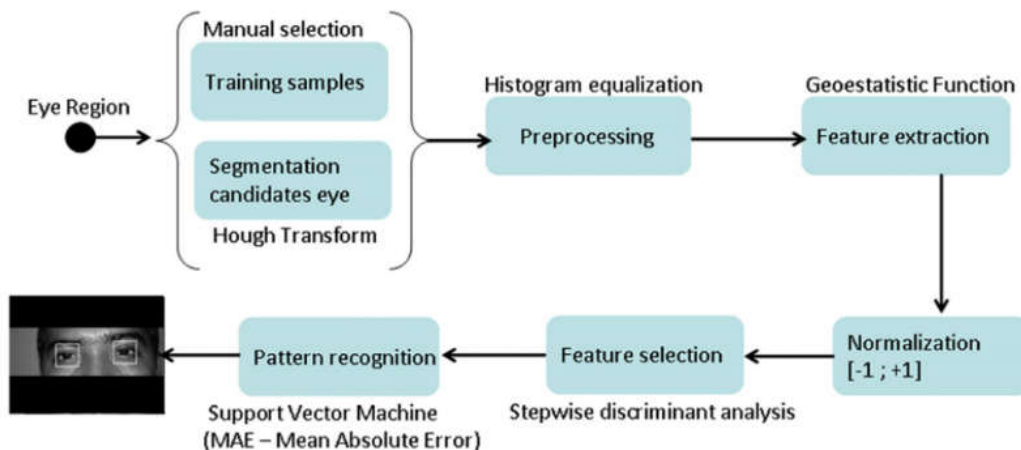
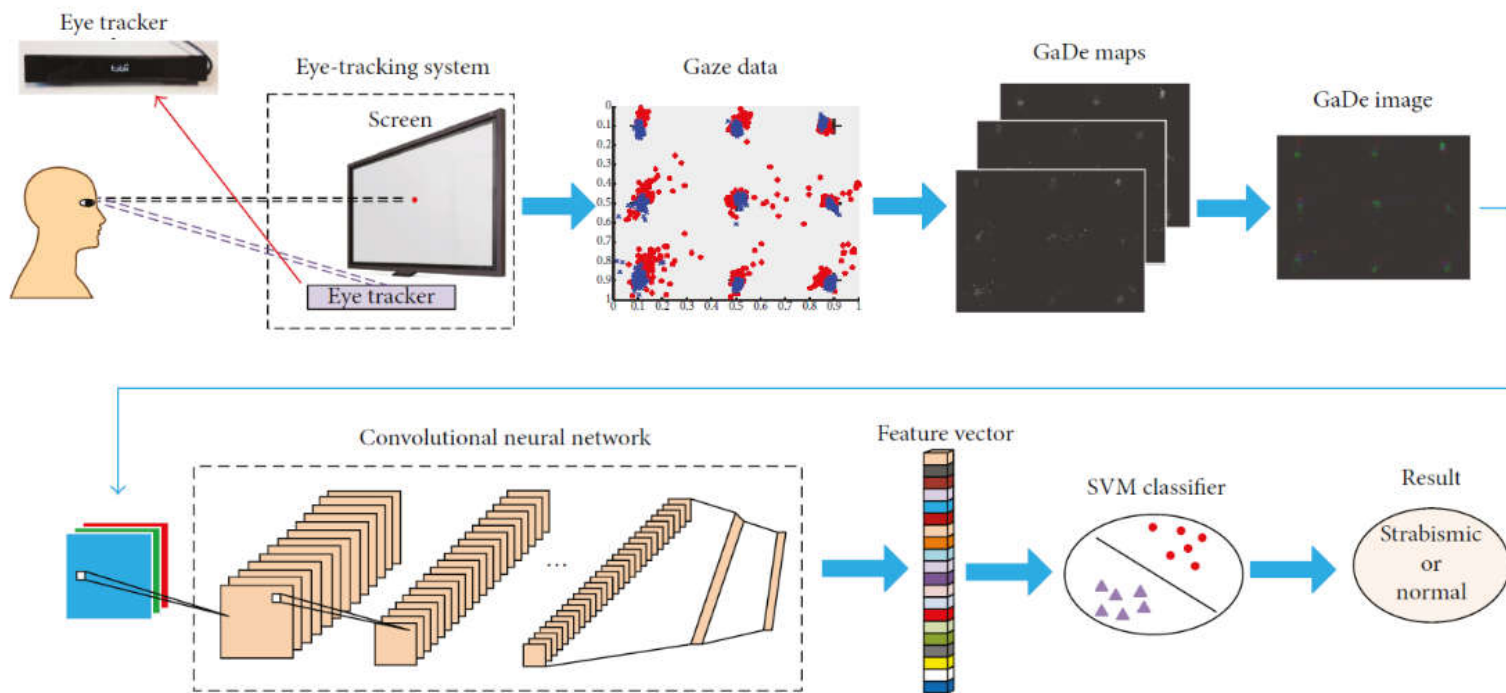


Fig. 4. Stages of the proposed methodology.



目前斜视检测的工作为数较少，上图的工作 **3** 使用传统图像处理方法结合机器学习方法进行斜视检测。

[4] De Almeida J O D S, Silva A C, De Paiva A C, et al. Computational methodology for automatic detection of strabismus in digital images through Hirschberg test[J]. Computers in biology and medicine, 2012, 42(1): 135-146.



上图的工作 **4** 则是使用眼动仪获取的数据结合卷积神经网络进行斜视检测，两者的识别准确率均不到**80%**。

[5] Chen Z, Fu H, Lo W L, et al. Strabismus Recognition Using Eye-Tracking Data and Convolutional Neural Networks[J]. Journal of Healthcare Engineering, 2018.

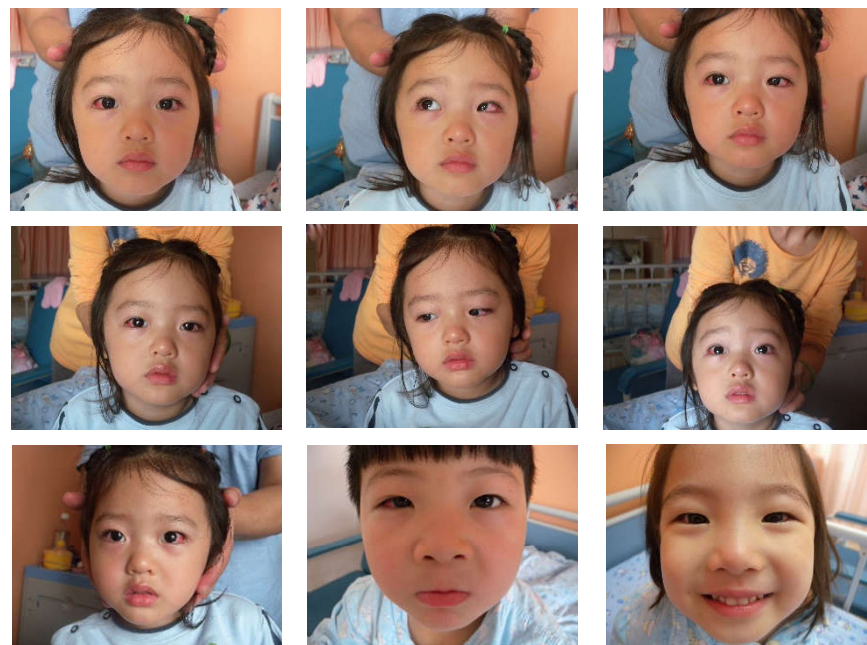
工作介绍



该数据集包括：13354张已被专业医生标注的从医院搜集的斜视图片，2943张已被标注的从医院搜集正常图片。



部分斜视图片



部分正常图片

算法流程图：

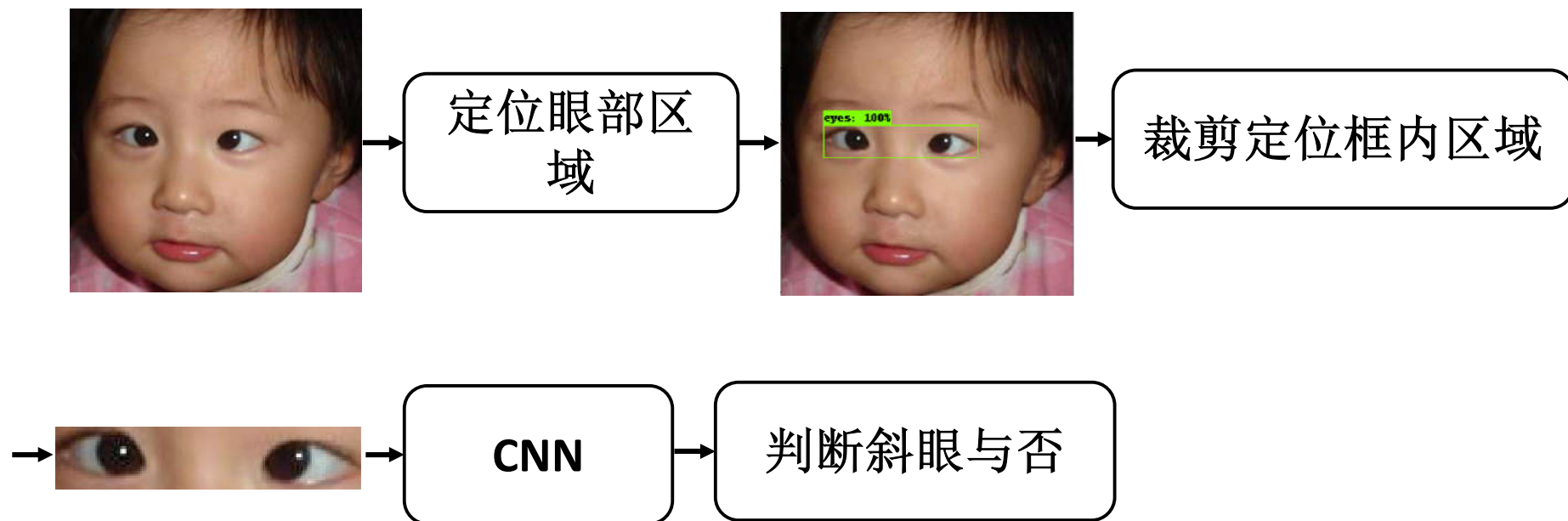
输入图片

定位眼部区域

裁剪定位框内区域

CNN

→ 判断斜眼与否

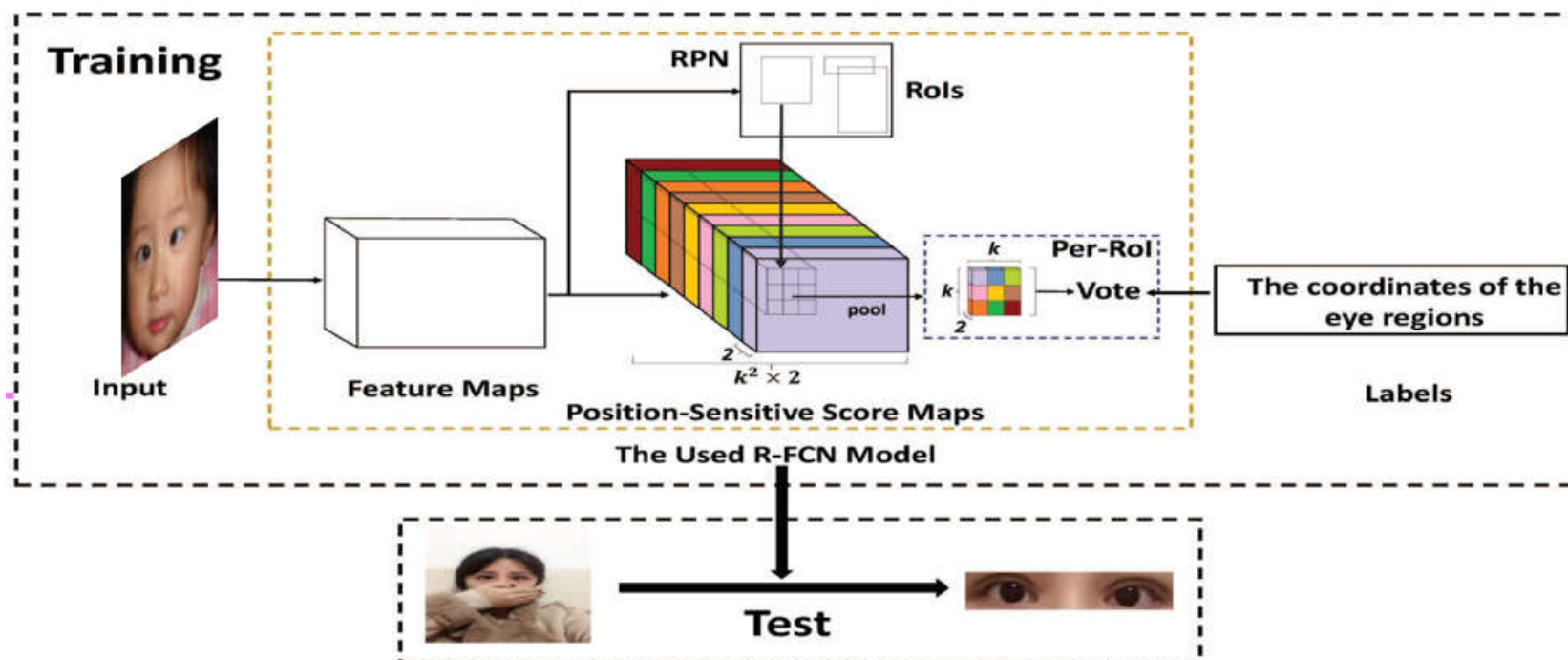


步骤1: 定位眼部区域

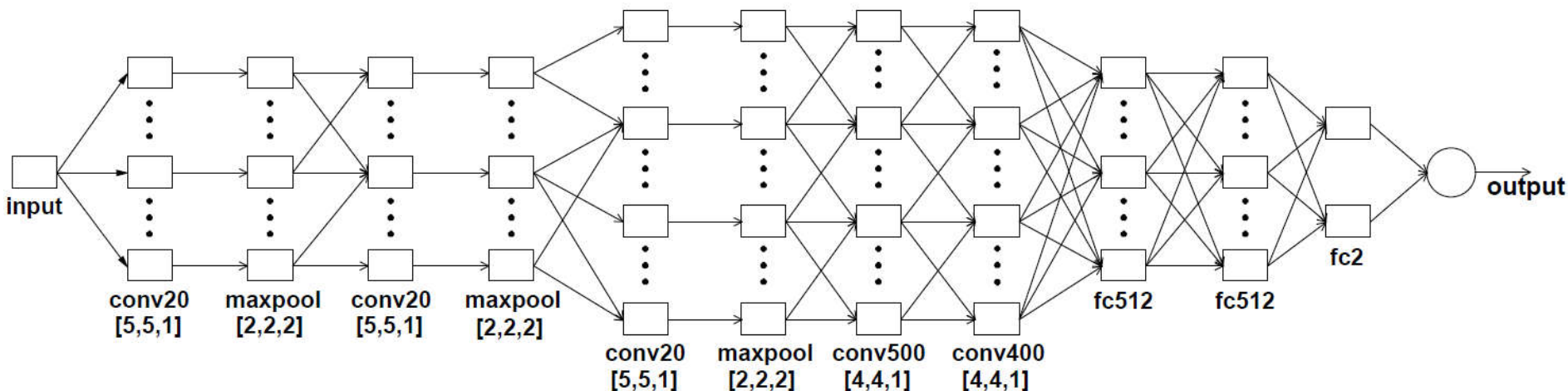
步骤2: 获取定位框内眼部区域

步骤3: 输入 进行斜眼识别

在上述算法即斜视检测系统中，眼部区域检测使用目标检测网络 7，网络结构如下图所示：



经过眼部区域检测网络得到眼部检测框后，进行裁剪得到眼部区域后输入下面卷积神经网络，该网络由5个卷积层、3个池化层、3个全连接层组成，最后输出斜眼识别结果。



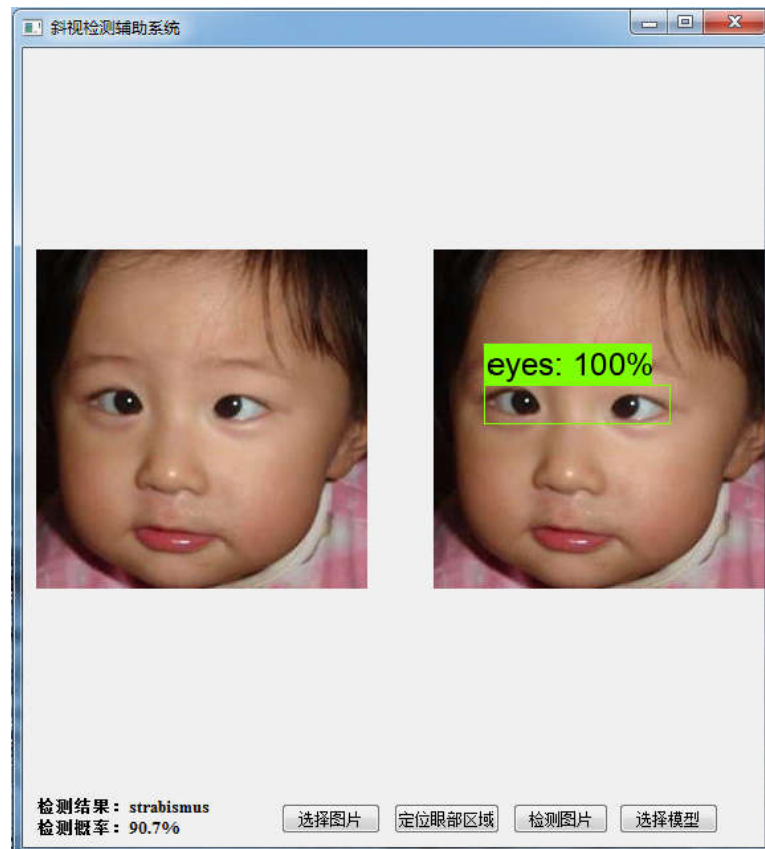
其中每层网络名称后跟的数字为和。

通道数，括号内容为

基于上述斜视检测系统 -
，将该算法封装开发斜
视检测界面软件如右图所示
，在实际使用过程中存在以
下缺点：

- 软件容量过大；
- 网络模型大，前向推导时
计算量大，占据电脑大量
内存空间。

因此，后续工作为对斜视检
测模型进行轻量化设计。



轻量级神经网络主要分为三种方法：

- 人工设计轻量级神经网络；
- 神经网络模型压缩算法；
- 基于神经网络架构搜索的自动化神经网络架构设计。

在该论文中采用的为第一种人工设计轻量级神经网络，其主要思想在于设计更高效的网络计算方式，主要是针对卷积的计算方法。

8 9 、 10
11 、 12 为其中的代表作。

[7] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

[8] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[9] Zhang, Xiangyu, et al. "Shufflenet: An extremely efficient convolutional neural network for mobile devices." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[10] Iandola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).

[11] Qin, Zheng, et al. "ThunderNet: Towards Real-Time Generic Object Detection on Mobile Devices." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.

系列提出深度可分离卷积
对标准的卷积进行分解。

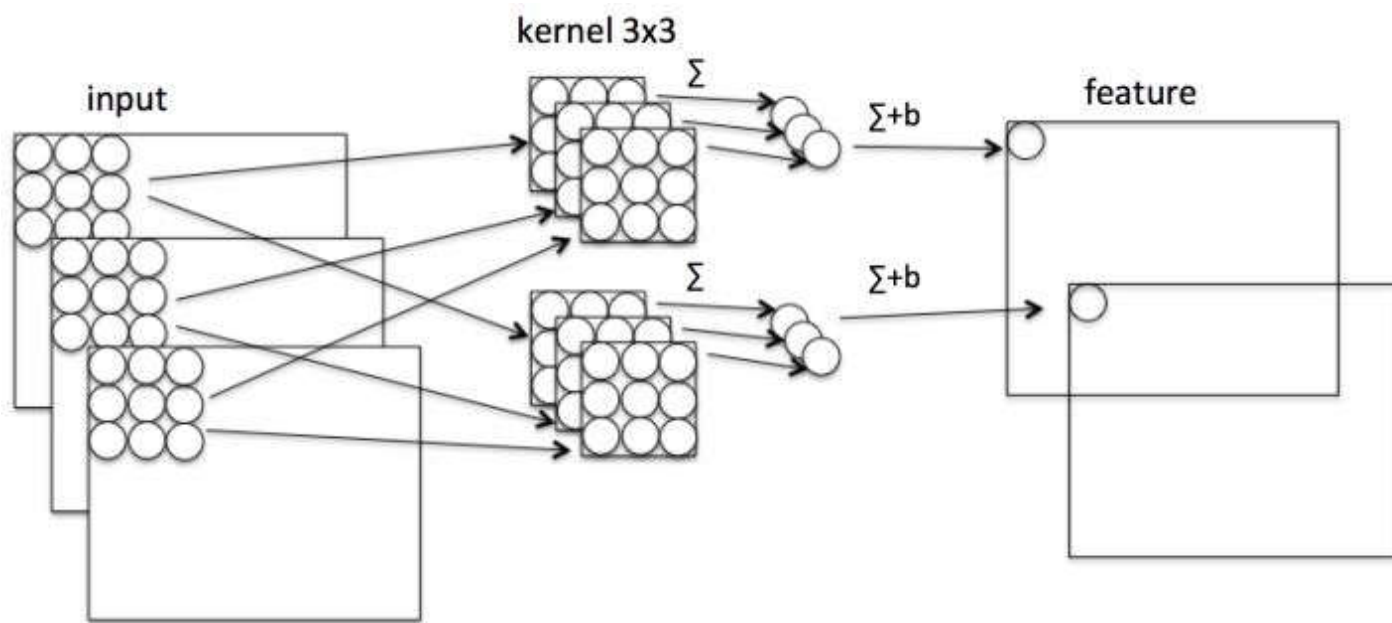
其中深度可分离卷积
分为：

- 深度卷积 ;
- 逐点卷积 。

可

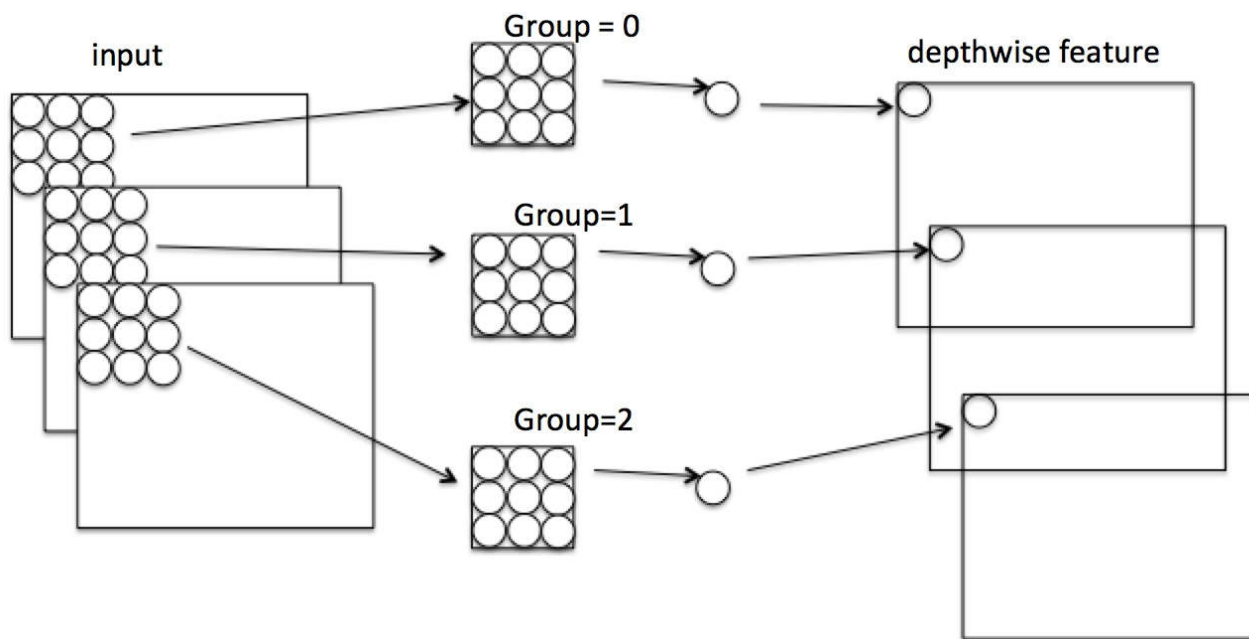
普通卷积操作：

- 假设有输入feature map，尺寸为 $H*W*C$ ，同时有 k 个 $3*3$ 卷积核，假设设置 $pad=1$ ， $stride=1$ ，输出尺寸为 $H'*W'*k$ 。



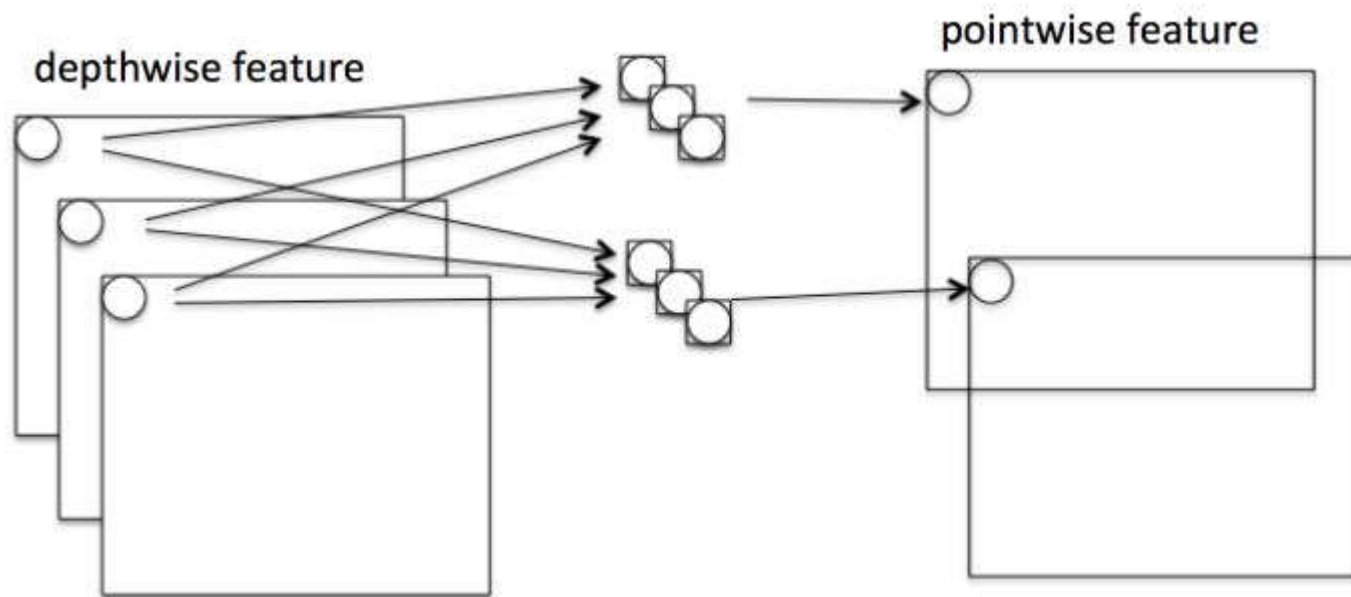
深度卷积操作：

- 深度卷积是指将 $H*W*C$ 的输入分为 C 组，然后每一组做 $3*3$ 的卷积，这样相当于收集了每个Channel的空间特征，即深度卷积特征。



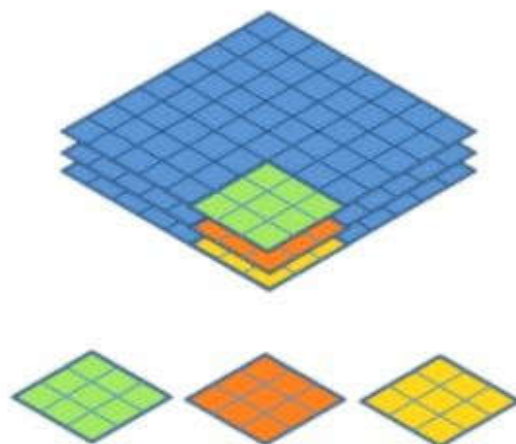
逐点卷积操作:

- 逐点卷积是指对 $H*W*C$ 的输入做 k 个普通 $1*1$ 的卷积, 这样相当于收集了每个点的特征, 即逐点卷积特征。深度卷积+逐点卷积最终输出也是 $H'*W'*k$ 。

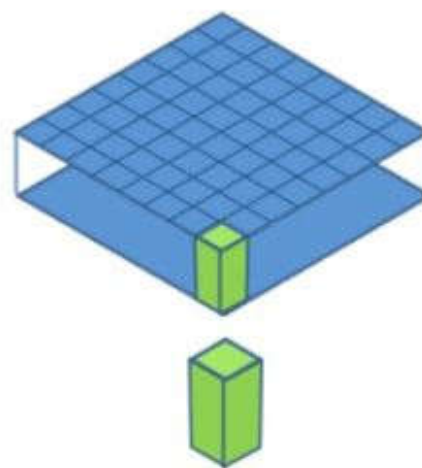


Mobilenet v1核心是把卷积拆分为深度卷积+逐点卷积两部分。

假设有 $H*W*C$ 的输入， k 个 $3*3$ 卷积核，将输入分为 C 组，每组做 $3*3$ 卷积，计算量为 $H*W*C*3*3$ 。再对输入做 k 个普通 $1*1$ 卷积，计算量为 $H*W*k$ 。而普通卷积计算量为 $H*W*C*k*3*3$ ，深度分离卷积计算量为普通卷积的 $1/k + 1/(3*3)$ 倍。

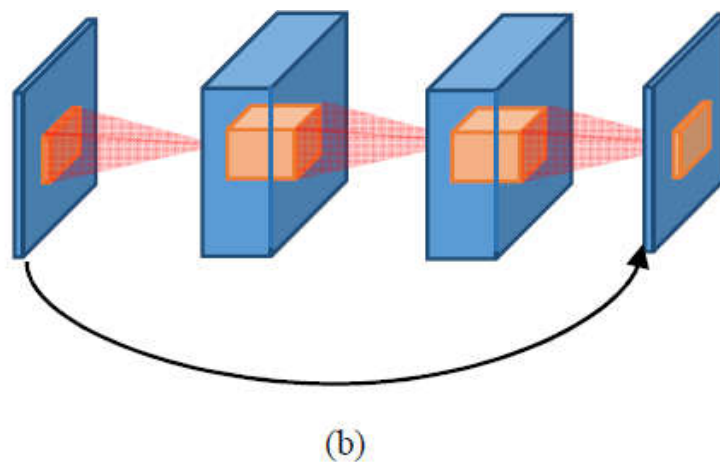
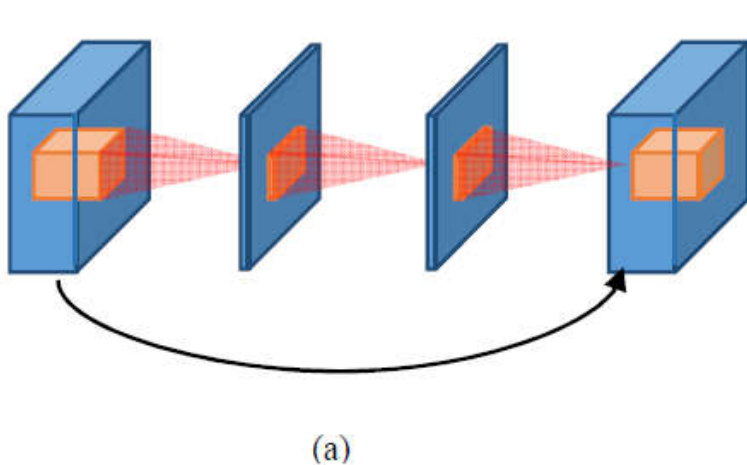


Depthwise Convolution



Pointwise Convolution

Mobilenet v2针对v1中ReLU造成的低维度数据坍塌和没有复用特征问题，提出了**Inverted residual block**作为网络基本结构，如下图所示，图a为普通残差网络结构，图b为反残差网络结构：



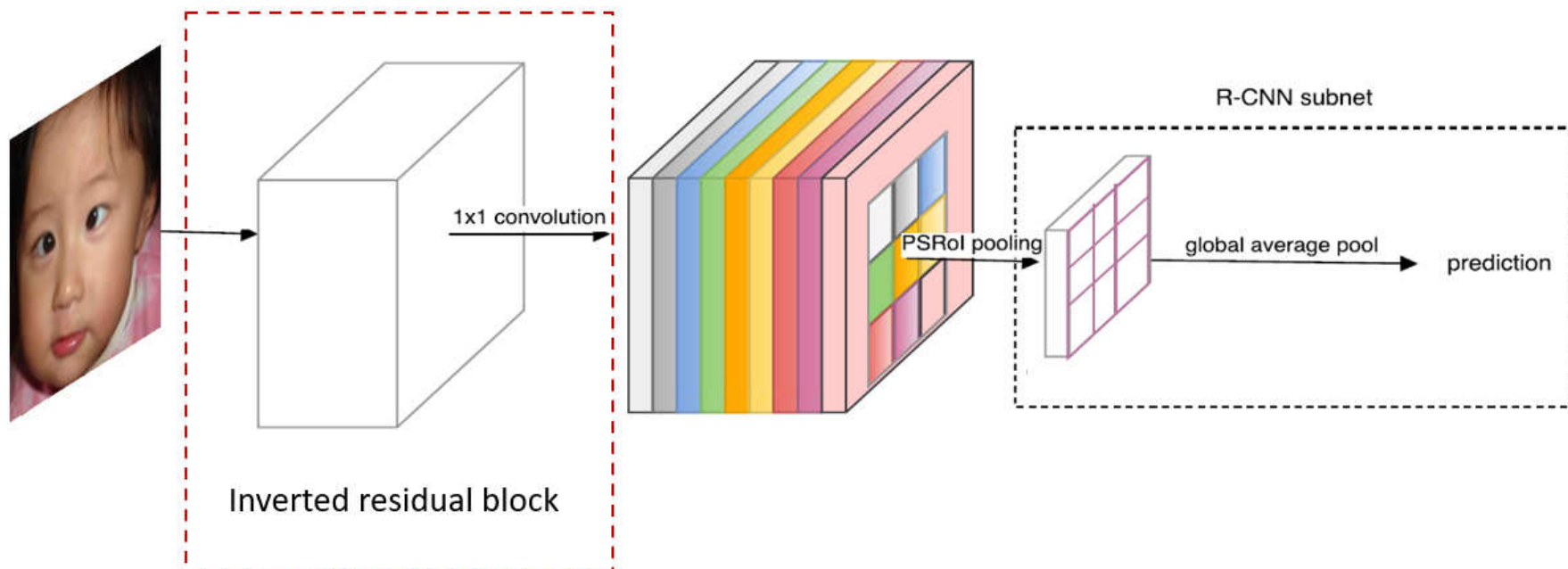


- 后续出的**EfficientNet**[13], **MixNet**[14]等轻量化网络几乎都是在深度分离卷积基础上的工作。
- 据此对斜视检测系统**RF-CNN**进行轻量化, 使用反残差结构代替**RF-CNN**中的普通卷积部分, 提出轻量化的**RF-CNN Lite**。

[12] Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *arXiv preprint arXiv:1905.11946* (2019).

[13] Tan, Mingxing, and Quoc V. Le. "Mixconv: Mixed depthwise convolutional kernels." *CoRR*, *abs/1907.09595* (2019).

在斜视检测系统 - 中的眼部区域检测网络部分，
即是 - 中，我们使用了上述的反残差结构取代了前
面提取 - 的部分网络，前期工作使用的为
-101网络。



由反残差结构组成的**ResNet-101**替代网络结构

Input	Operator	t	c	n	s
256*256*3	conv2d	-	32	1	2
128*128*32	bottleneck	1	16	1	2
64*64*16	bottleneck	6	32	2	2
32*32*32	bottleneck	6	64	3	2
16*16*64	bottleneck	6	128	4	2
8*8*128	bottleneck	6	256	3	1
8*8*256	Conv2d1*1	-	1024	1	1

其中 为反残差结构中升维的膨胀系数， 为输出通道数， 为重复次数， 为步长。

在斜视检测系统中斜眼识别部分中，前期工作使用的是手工搭建普通卷积的神经网络，为了使其更加轻量化，使用反残差结构进行代替，网络结构如下图所示：

斜眼识别替代网络结构

Input	Operator	t	c	n	s
224*224*3	conv2d	-	32	1	2
112*112*32	bottleneck	1	16	1	2
56*56*16	bottleneck	6	32	2	2
28*28*32	bottleneck	6	64	3	2
14*14*64	bottleneck	6	128	4	2
7*7*128	Conv2d1*1	-	1024	1	1
7*7*1024	Avgpool7*7	-	-	1	-
1*1*1024	Conv2d1*1	-	2	-	-

其中 t 为反残差结构中升维的膨胀系数， c 为输出通道数， n 为重复次数， s 为步长。

R-FCN不同版本对比实验结果

Network	Average Precision	Param memory	flops	time/img(Titan Xp)
R-FCN + ResNet50	86.52%	122MB	79 GFLOPS	0.12sec
R-FCN + ResNet101	88.27%	194MB	117 GFLOPS	0.14sec
R-FCN + Inverted residual block	84.64%	22MB	2 GFLOPS	0.09sec

该实验结果均在斜眼数据集中进行实验，输入图片尺寸为**256 256**，其中
为平均精度，为参数所占内存，为浮点运算数
，time/img为每张图片检测所花费时间。

斜眼识别CNN不同版本对比实验结果

Network	Param mem	flops	TP	TN	FP	FN	Se	Sp	Acc	AUC
CNN(普通卷积层)	51MB	2GFLOPs	1085	1146	78	68	0.9329	0.9440	0.9386	0.9854
CNN(反残差结构)	16MB	579 MFLOPs	1012	1161	115	89	0.8980	0.9288	0.9163	0.9742

测试集总共为**2377**张图片。其中 为参数所占内存， 为浮点运算数，TP为真正例，TN为真反例，FP为假正例，FN为假反例，Se为Sensitive敏感度，Sp为Specificity特效度，Acc为Accuracy准确率，AUC为Area Under Curve（ROC曲线下的面积）。

创新点总结:

- 本论文针对斜视检测实现远程医疗，与当前流行的深度学习算法联系起来，提出了斜视检测系统RF-CNN，同时建立了斜视数据集并开发了供眼科医生使用的斜视检测软件；
- 针对软件使用过程中体积大的缺点对该斜视检测系统进行轻量化设计，利用反残差结构提出了轻量化的斜视检测系统RF-CNN Lite，在准确率以及精确率等指标下降2%左右的情况下，可以达到参数量减少三分之二，计算量减少三分之二的效果，对远程医疗的软件进行有效精简，得到了体积更小、速度更快的斜视检测系统，实现了轻便性。

不足之处:

- 本论文采用的斜视训练数据集针对去医院就诊的患者，对其他应用场景的数据的泛化性能没有达到医院临床使用的标准。
- 在针对斜视检测模型的轻量化设计工作中，只使用了人工设计轻量级神经网络方式对模型进行轻量化设计，没有使用其他的神经网络轻量化方式如基于神经网络架构搜索的自动化神经网络架构设计、神经网络模型压缩算法进行模型效果对比。

总结与展望:

- 针对斜视检测的对其他场景的泛化性能对训练数据可以采用更加丰富多样的图像以拓宽模型的应用场景，并采用更多的数据增广方式，对于一些特殊场景丰富数据集可以进一步提高模型的准确率。
- 可以使用其他的神经网络轻量化方式验证模型效果，与当前的方法进行对比，用以选取更合适的网络模型轻量化方法。

- [1] “Visual Processing: Strabismus”. National Eye Institute. National Institutes of Health. June 16, 2010. Archived from the original on October 5, 2016. Retrieved October 2, 2016.
- [2] Kiorpes, L., Kiper, D. C., O’ keefe, L. P., Cavanaugh, J. R. & Movshon, J. A. Neuronal correlates of amblyopia in the visual cortex of macaque monkeys with experimental strabismus and anisometropia. *J. Neurosci.* 18, 6411–6424 (1998)
- [3] Van de Graaf, E. et al. Amblyopia and strabismus questionnaire: design and initial validation. *Strabismus* 12, 181–193 (2004).
- [4] De Almeida J O D S, Silva A C, De Paiva A C, et al. Computational methodology for automatic detection of strabismus in digital images through Hirschberg test[J]. *Computers in biology and medicine*, 2012, 42(1): 135-146.
- [5] Chen Z, Fu H, Lo W L, et al. Strabismus Recognition Using Eye-Tracking Data and Convolutional Neural Networks[J]. *Journal of Healthcare Engineering*, 2018.
- [6] Dai, Jifeng, et al. "R-fcn: Object detection via region-based fully convolutional networks." *Advances in neural information processing systems*. 2016.
- [7] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

- [8] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [9] Zhang, Xiangyu, et al. "Shufflenet: An extremely efficient convolutional neural network for mobile devices." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [10] Iandola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).
- [11] Qin, Zheng, et al. "ThunderNet: Towards Real-Time Generic Object Detection on Mobile Devices." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.
- [12] Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *arXiv preprint arXiv:1905.11946* (2019).
- [13] Tan, Mingxing, and Quoc V. Le. "Mixconv: Mixed depthwise convolutional kernels." *CoRR, abs/1907.09595* (2019).

欢迎各位专家提出宝贵意见！