

# 基于深度学习的图像质量增强方法研究

黄昱博

北京邮电大学

huangyubo@bupt.edu.cn

硕士学位论文答辩

2023 年 8 月 20 日

## ① 研究背景

## ② 研究内容

### 基于深度级注意力机制的虹膜图像修复方法

理论分析

实验配置与结果

### 基于生成式先验知识的虹膜图像修复方法

理论分析

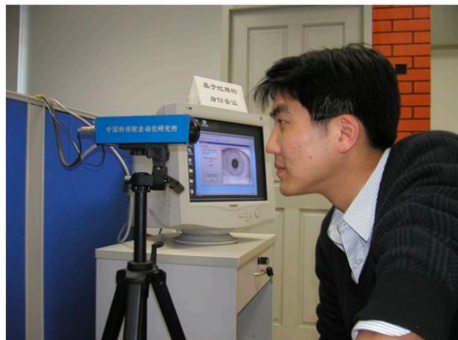
实验配置与结果

## ③ 总结与展望

# 研究背景

图像质量增强技术是指将低质量图像修复成高质量图像的技术,被广泛应用于生物特征识别领域。

其中,虹膜识别技术是一种非接触式、高安全性、高可靠性的身份认证技术,是目前最有前景的生物特征识别技术之一。



# 研究背景

虹膜识别的瓶颈问题之一是难以收集到高质量的虹膜图像。虹膜图像通常伴随以下三种退化：

- 在远距离场景下，采集到的虹膜图像**分辨率太低**。
- 如果受试者站在相机的景深之外，会产生**离焦模糊**。
- 在采集过程中，如果受试者在移动，会造成**运动模糊**。



(a) 原图



(b) 低分辨率



(c) 离焦模糊



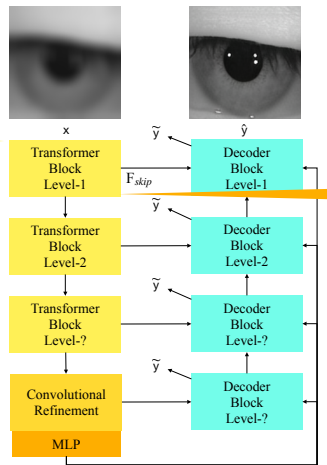
(d) 运动模糊

# 基于深度级注意力机制的虹膜图像修复方法

## 骨干网络

现有的方法一般运用了卷积神经网络，只能利用局部信息，无法利用全局信息，导致特征提取能力较弱，无法精确地修复虹膜图像。

本章提出了基于深度级注意力机制的虹膜图像修复方法，可以对长距离的像素依赖进行建模，并且能在推理过程中，根据输入来动态地调整参数，同时，实现了线性的计算复杂度，使得本章方法的计算量小于现有的基于注意力机制的方法。



# 基于深度级注意力机制的虹膜图像修复方法

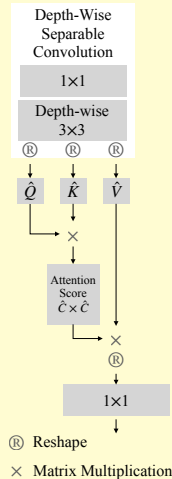
## 深度级注意力机制

在传统的自注意力机制中，注意力分数  $\alpha = Q \cdot K$  的时间和空间复杂度为  $\mathcal{O}(W^2H^2)$ 。因此，将这样的自注意力机制用到图像修复任务中，会产生过大的计算开销。

为了缓解这个问题，本章提出了具有线性复杂度的深度级注意力机制。

- ① 使用像素卷积，收集一个像素在所有的通道范围内的内容信息，这个结构可以进一步提取通道范围内的特征。
- ② 使用深度卷积，提取一个通道中，所有空间范围内的内容信息。

(a) 通道注意力机制



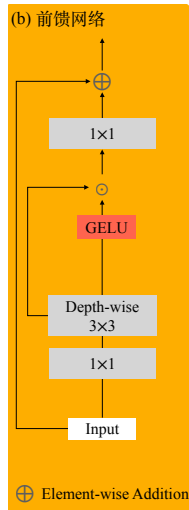
# 基于深度级注意力机制的虹膜图像修复方法

## 深度级激活网络

传统的激活网络以完全相同的方法独立地处理每个像素上的信息。

本章采用了深度级激活网络。该网络中的两个机制，对提升表征学习能力有所帮助：

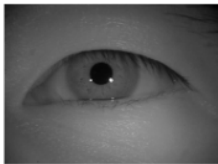
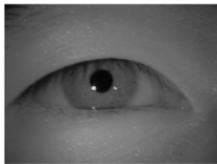
- ① 门控机制：输入并行经过两条线性变换路径，其中一条路径添加了 GELU 非线性激活函数，然后将这两条路径做数量积。
- ② 深度级卷积：保持了从相邻像素编码信息的能力，此外，可以有效地学习到图像局部的结构。



## 实验数据

本文使用了中科院虹膜数据集 CASIA-IrisV4 中的

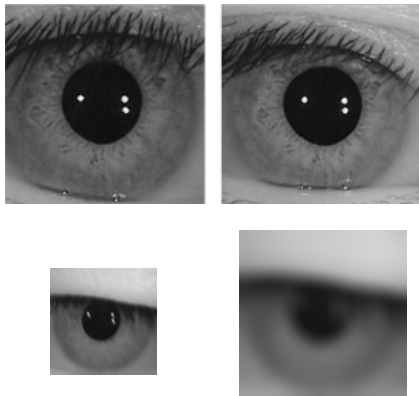
- ① 灯光数据集 CASIA-Iris-Lamp，包含来自 411 个受试者的 16,212 张虹膜图像，共 822 个类别。类内变化的主要来源是瞳孔的收缩。
- ② 千人数据集 CASIA-Iris-Thousand，包含来自 1,000 个受试者的 20,000 张虹膜图像，共 2,000 个类别。类内变化的主要来源是眼镜的镜面反射。





# 实验数据

- 本文先将上述两数据集进行合并，删除虹膜处有遮挡的数据，得到 28,089 张虹膜图像，其中包含 2,738 个类别，用于本文的实验。
- 接着用虹膜检测算法定位虹膜，然后进行裁剪，分辨率由  $640 \times 480$  变为  $224 \times 224$ 。
- 最后将图像通过随机的高斯模糊核、运动模糊核和下采样因子进行退化，用来模拟真实世界中图像的退化。
- 测试集与训练集的合成方式相同。并且，二者没有重叠。



# 基于深度级注意力机制的虹膜图像修复方法

## 实验设置

- Transformer 模块的层级数为 4，每个层级包含 1 个 Transformer 块。从层级 1 到层级 4，注意力机制“头”的数目分别为  $[1, 2, 4, 8]$ 。
- 由浅到深每个层级输出图像分辨率分别为  $[256^2, 128^2, 64^2, 32^2, 16^2, 8^2, 4^2]$ 。
- Adam 优化器，学习率设为  $2 \times 10^{-4}$ 。
- 深度学习框架为 PyTorch。
- GPU 型号为 NVIDIA Tesla v100。

 PyTorch



# 定量实验结果

Method	AUC $\uparrow$	EER $\downarrow$	TAR@FAR =0.001 $\uparrow$	TAR@FAR =0.01 $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
输入	0.9521	0.1153	0.1992	0.5361	30.68	0.8336	0.2865
EDSR	0.9796	0.0705	0.5356	0.7538	34.15	0.8808	0.1249
RCAN	0.9818	0.0679	0.5608	0.7666	33.57	0.8789	0.1324
SRGAN	0.9732	0.0853	0.3960	0.6705	32.85	0.8695	0.1311
ESRGAN	0.9585	0.1093	0.4816	0.6876	30.29	0.8314	0.1189
SISN	0.9836	0.0542	0.5878	0.8274	32.36	0.8729	0.1350
本章方法	0.9864	0.0532	0.7475	0.8773	32.82	0.8754	0.1513
原始图像	0.9914	0.0274	0.9323	0.9613	-	-	-

表: 基于深度级注意力机制的虹膜图像修复方法与现有方法的定量结果对比。

# 定性实验结果

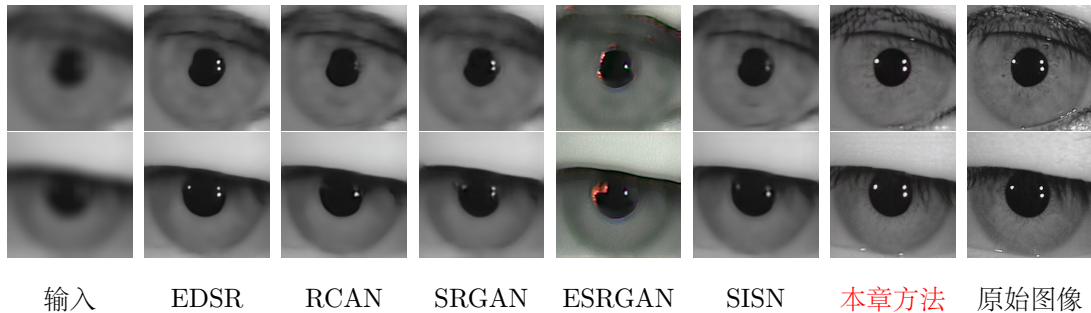


图: 基于深度级注意力机制的虹膜图像修复方法与现有方法的定性实验结果对比。

# 虹膜识别性能测试

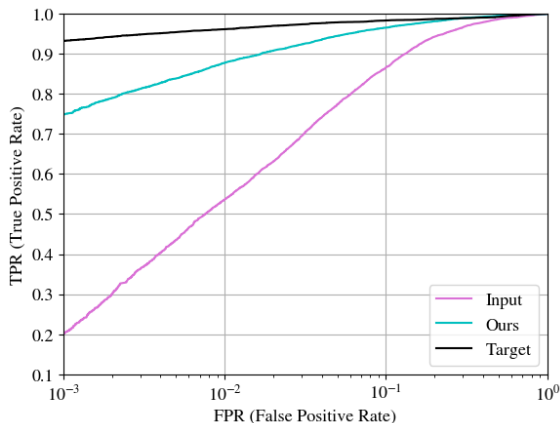


图: 基于深度级注意力机制的虹膜图像修复方法的 ROC 曲线。

Method	AUC $\uparrow$	EER $\downarrow$	TAR@FAR =0.001 $\uparrow$	TAR@FAR =0.01 $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
w/o Transformer	0.9523	0.1251	0.5342	0.6753	30.01	0.8072
w/o Pyramid Loss	0.9905	0.0456	0.7908	0.9033	33.42	0.8791
w/o Perceptual Loss	0.9521	0.1252	0.5266	0.6725	29.50	0.8068
w/o Adversarial Loss	0.9864	0.0532	0.7200	0.8702	32.12	0.8629
本章方法	0.9864	0.0532	0.7475	0.8773	32.82	0.8754

表: 基于深度级注意力机制的虹膜图像修复方法的消融实验结果。

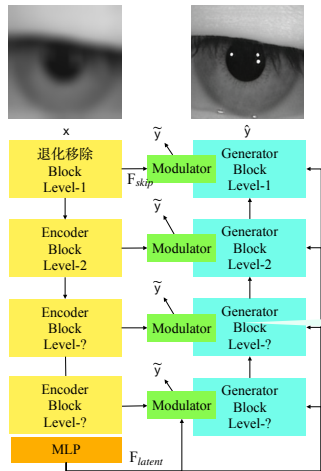
# 基于生成式先验知识的虹膜图像修复方法

## 骨干网络

生成对抗网络一般用于直接生成目标图像。由于虹膜具有复杂的微结构，借助先验知识有助于修复得到更加逼真的纹理细节。

虹膜生成式先验知识是一个预训练好的虹膜生成对抗网络，它捕获了虹膜的特征分布，储存到了网络的参数中。本章利用虹膜生成式先验知识，指导图像修复工作。

解码器为 StyleGANv2 的生成器，包含了多层级的生成器块，生成器块的数目由输入图像的分辨率决定。



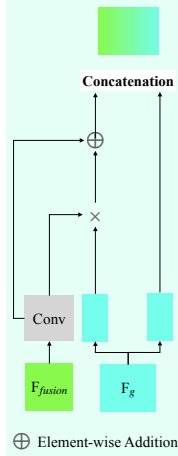
# 基于生成式先验知识的虹膜图像修复方法

## 虹膜特征调制模块

为了嵌入虹膜生成式先验知识，本章设计了虹膜特征调制模块：

- ① 将编码器的两个输出（跳跃特征图和隐向量）进行融合。
- ② 将生成器输出的特征在通道维度上平均分成两部分，一部分保持不变，另一部分进行调制，最后把这两部分连接在一起，就得到了最终的特征。

通道分离特征转换模块





# 基于生成式先验知识的虹膜图像修复方法

## 实验设置

- 实验数据的处理与上一章相同。
- 采用的多层级网络结构的层级数为 7，由浅到深每个层级输出图像分辨率分别为  $[256^2, 128^2, 64^2, 32^2, 16^2, 8^2, 4^2]$ 。
- Adam 优化器，学习率设为  $2 \times 10^{-3}$ 。
- 深度学习框架为 PyTorch。
- GPU 型号为 NVIDIA Tesla v100。

 PyTorch



# 定量实验结果

Method	AUC $\uparrow$	EER $\downarrow$	TAR@FAR =0.001 $\uparrow$	TAR@FAR =0.01 $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
输入	0.9521	0.1153	0.1992	0.5361	30.68	0.8336	0.2865
EDSR	0.9796	0.0705	0.5356	0.7538	34.15	0.8808	0.1249
RCAN	0.9818	0.0679	0.5608	0.7666	33.57	0.8789	0.1324
SRGAN	0.9732	0.0853	0.3960	0.6705	32.85	0.8695	0.1311
ESRGAN	0.9585	0.1093	0.4816	0.6876	30.29	0.8314	0.1189
SISN	0.9836	0.0542	0.5878	0.8274	32.36	0.8729	0.1350
本章方法	<b>0.9883</b>	<b>0.0505</b>	<b>0.7843</b>	<b>0.8873</b>	32.88	0.8631	0.1472
原始图像	0.9914	0.0274	0.9323	0.9613	-	-	-

表: 基于生成式先验知识的虹膜图像修复方法与现有方法的定量结果对比。

# 定性实验结果

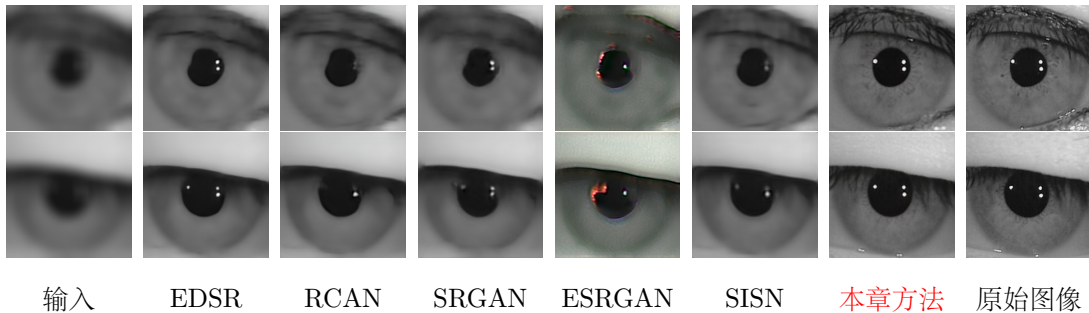


图: 基于生成式先验知识的虹膜图像修复方法与现有方法的定性实验结果对比。

# 虹膜识别性能测试

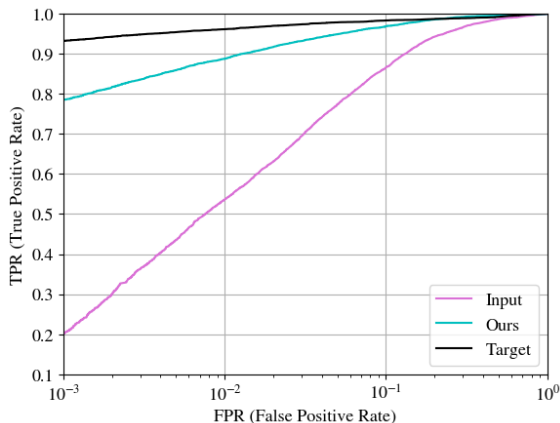


图: 基于生成式先验知识的虹膜图像修复方法的 ROC 曲线。

Method	AUC $\uparrow$	EER $\downarrow$	TAR@FAR =0.001 $\uparrow$	TAR@FAR =0.01 $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
w/o Generative Iris Prior	0.9513	0.1273	0.5145	0.6678	29.43	0.8023
w/o Chanel-Split	0.9903	0.0435	0.7889	0.9060	32.96	0.8734
w/o SFT	0.9879	0.0480	0.7721	0.8917	32.12	0.8582
本章方法	0.9883	0.0505	0.7843	0.8873	32.88	0.8631

表: 基于生成式先验知识的虹膜图像修复方法的消融实验结果。

# 工作总结

传统自注意力机制的计算复杂度随着图片分辨率的增大而以平方的速度增长。这导致 Transformer 无法有效地应用到高分辨率的图像修复任务中。本文采用了作用于通道维度的自注意力机制，它拥有线性的计算复杂度，使得 Transformer 在图像修复任务中的应用变得更加切实可行。

传统的激活网络以完全相同的方法独立地处理每个像素上的信息。本文设计了深度级激活网络，不仅具有从相邻像素编码信息的能力，还可以有效地学习到图像局部的结构，丰富了图像中的细节。

传统的图像质量增强方法使用生成对抗网络直接生成最终图像，没有利用图像的特性。本文设计了虹膜特征调制模块，嵌入了虹膜生成式先验知识，使得输出图像的纹理更加真实。

# 工作展望

本文的方法中，如果去掉金字塔损失，识别性能上升，图像质量下降。金字塔损失由 L1 损失构成。后续若想进一步提升识别性能，需减少对 L1 损失的使用。

本文的方法中，如果不使用通道分离特征转：不是原始方法中对一半通道进行特征转换，而是对所有通道进行特征转换，会使计算复杂度增大，识别性能和图像质量上升。后续若想要进一步提升识别性能，可以将所有通道进行特征转换。

本文的方法参数量较大，适合部署在算力较强的平台上。近年来，为了提升计算效率，具有多种处理器或内核（如 CPU、GPU、ASIC、FGA 和 NPU）的系统被越来越广泛地使用。为了适应这一趋势，基于深度学习的图像质量增强方法还需要增强其可扩展性。例如，根据不同平台的算力、不同的数据、每张图像中的不同区域，智能地调节方法中使用到的网络层级数目，从而在性能与效率之间寻求最优平衡。

Method	#Parameters ↓	Inference Time ↓
EDSR	43M	88ms
RCAN	16M	70ms
SRGAN	1M	10ms
ESRGAN	17M	154ms
DeblurGANv2	11M	<u>15ms</u>
HiFaceGAN	161M	91ms
SISN	<u>8M</u>	202ms
SparNet	11M	134ms
本文的两种方法	70M	150ms

表: 本文的方法与现有方法参数量和推理时间的对比。



# 攻读学位期间发表的学术论文

- [1] **Huang Yubo**, Wang Jia, Li Peipei, Xiang Liuyu, Li Peigang, He Zhaofeng. Generative Iris Prior Embedded Transformer for Iris Restoration [A]. // 2023 IEEE International Conference on Multimedia and Expo (ICME) [C]. Los Alamitos: IEEE Computer Society, 2023: 1–6. (**CCF-B**, in press).

# 谢谢

欢迎各位老师指正！