# 一键卸妆与视频超分辨率：腾讯优图ICCV 2017分享

沈小勇
腾讯优图
goodshenxy@gmail.com

# About Me (http://Xiaoyongshen.me)

- BS, MS from ZJU and PhD from CUHK
  - Supervisors: Prof. Ligang and Prof. Jiaya
  - Three years research in CG and five years in CV
- Senior researcher in Youtu
  - Lead the research group
- My research is mainly on
  - Image filtering and restoration, matting, deblur, etc.
  - Motion and depth estimation, segmentation
  - Image classification, object detection and semantic segmentation, etc.

腾讯优图

# 1 关于优图实验室-概览

优图团队立足于社交网络大平台，借助社交业务积累的海量人脸、图片、音乐等数据，专注在人脸、图像、音乐、语音、机器学习等领域开展技术研究，并积极推动研究成果在业务中落地产生价值。

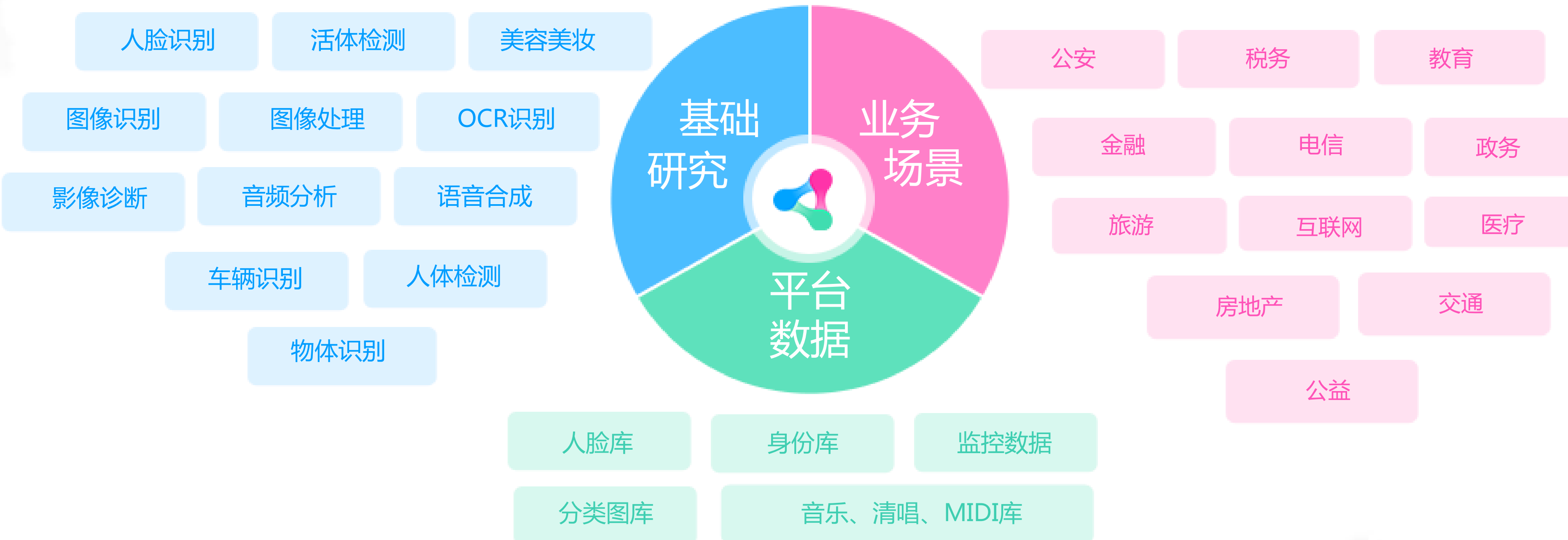
人脸识别


图像识别


音频识别

- 2017年4月，在国际MegaFace海量人脸识别数据库刷新世界记录
- 2017年3月，在国际LFW人脸数据库上刷新世界纪录
- 2017年3月，在国际ICDAR 2015文本检测项目中刷新世界纪录
- 2016年，优图实验室获得“腾讯行业贡献奖”
- 2015年，优图哼唱识别技术获得“腾讯年度微创新奖”
- 2015年10月，在国际音频比赛MIREX的哼唱识别比赛中，取得总成绩世界的第一，并刷新其中一项世界纪录
- 2015年4月，在国际Pascal VOC2012物体分类赛刷新世界纪录
- 2014年11月，腾讯优图人脸检测刷新FDDB世界纪录
- 2014年，人脸识别获得腾讯公司年度“重大技术突破奖”
- 2014年，联合上海交通大学获得上海市科技进步二等奖
- 2013年，优图压缩获得腾讯公司年度“重大技术突破奖”
- 2012年，优图电商联合团队获得腾讯公司级别“卓越运营奖”

# 1 关于优图实验室-概览




**研究、场景、数据三者融合**

# 1 关于优图实验室-最新技术突破

## 2017年3月

在**ICDAR 2015**比赛中刷新Focused Scene Text挑战的Text Localization项目世界纪录

| Deteval | | | |
|---|---|---|---|
| Method | Recall | Precision | Hmean |
| Tencent Youtu | 89.53 % | 94.26 % | 91.84 % |
| CNN based mo... | 89.17 % | 94.63 % | 91.82 % |
| RRPN-4 | 87.31 % | 95.19 % | 91.08 % |
| MSRA_v1 | 88.58 % | 93.67 % | 91.06 % |
| SRC-B-Machine... | 87.07 % | 93.28 % | 90.07 % |
| Baidu IDL | 87.11 % | 92.83 % | 89.88 % |
| CAS_HotEye_ver2 | 84.31 % | 94.17 % | 88.97 % |
| XvBaoBao | 85.37 % | 91.46 % | 88.31 % |
| CAS_HotEye | 86.65 % | 89.99 % | 88.29 % |

## 2017年3月

在**LFW**无限制条件下人脸验证测试中，优图提交的最新成绩为**99.80%**，提升了上次99.65%的成绩，再次在这一测试中刷新纪录。

| | |
|---|---|
| Faceall[71] | 0.9940 ± 0.0010 |
| JustMeTalk[72] | 0.9887 ± 0.0016 |
| Facevisa[74] | 0.9955 ± 0.0014 |
| pose+shape+expression augmentation[75] | 0.9807 ± 0.0060 |
| ColorReco[76] | 0.9940 ± 0.0022 |
| Asaphus[77] | 0.9815 ± 0.0039 |
| Daream[78] | 0.9968 ± 0.0009 |
| Dahua-FaceImage[80] | 0.9978 ± 0.0007 |
| Easen Electron[81] | 0.9968 ± 0.0009 |
| Skytop Gaia[82] | 0.9630 ± 0.0023 |
| CNN-3DMM estimation[83] | 0.9235 ± 0.0129 |
| Samtech Facequest[84] | 0.9971 ± 0.0018 |
| XYZ Robot[87] | 0.9895 ± 0.0020 |
| THU CV-AI Lab[88] | 0.9973 ± 0.0008 |
| PingAn Tech[89] | 0.9960 ± 0.0031 |
| dlib[90] | 0.9938 ± 0.0027 |
| Aureus[91] | 0.9920 ± 0.0030 |
| YouTu Lab, Tencent[63] | 0.9980 ± 0.0023 |

Table 6: Mean classification accuracy $\hat{\mu}$ and standard error of the mean $S_E$.

## 2017年4月

在国际权威海量人脸识别数据库**MegaFace**中，以**83.290%**的最新成绩在100万级别人脸识别测试中拔得头筹

Rank-1 Identification Accuracy with 1 Million Distractors

| Algorithm | Date Submitted | Set 1 |
|---|---|---|
| YouTu Lab (Tencent Best-Image) | 04/08/2017 | 83.290% |
| DeepSense V2 | 1/22/2017 | 81.298% |
| Vocord-deepVo1.2 | 12/1/2016 | 80.258% |
| GRCCV | 12/1/2016 | 77.677% |
| SphereFace - Small | 12/1/2016 | 75.766% |

# 优图与ICCV

## 计算机视觉顶级会议ICCV 2017 腾讯优图入选12篇论文

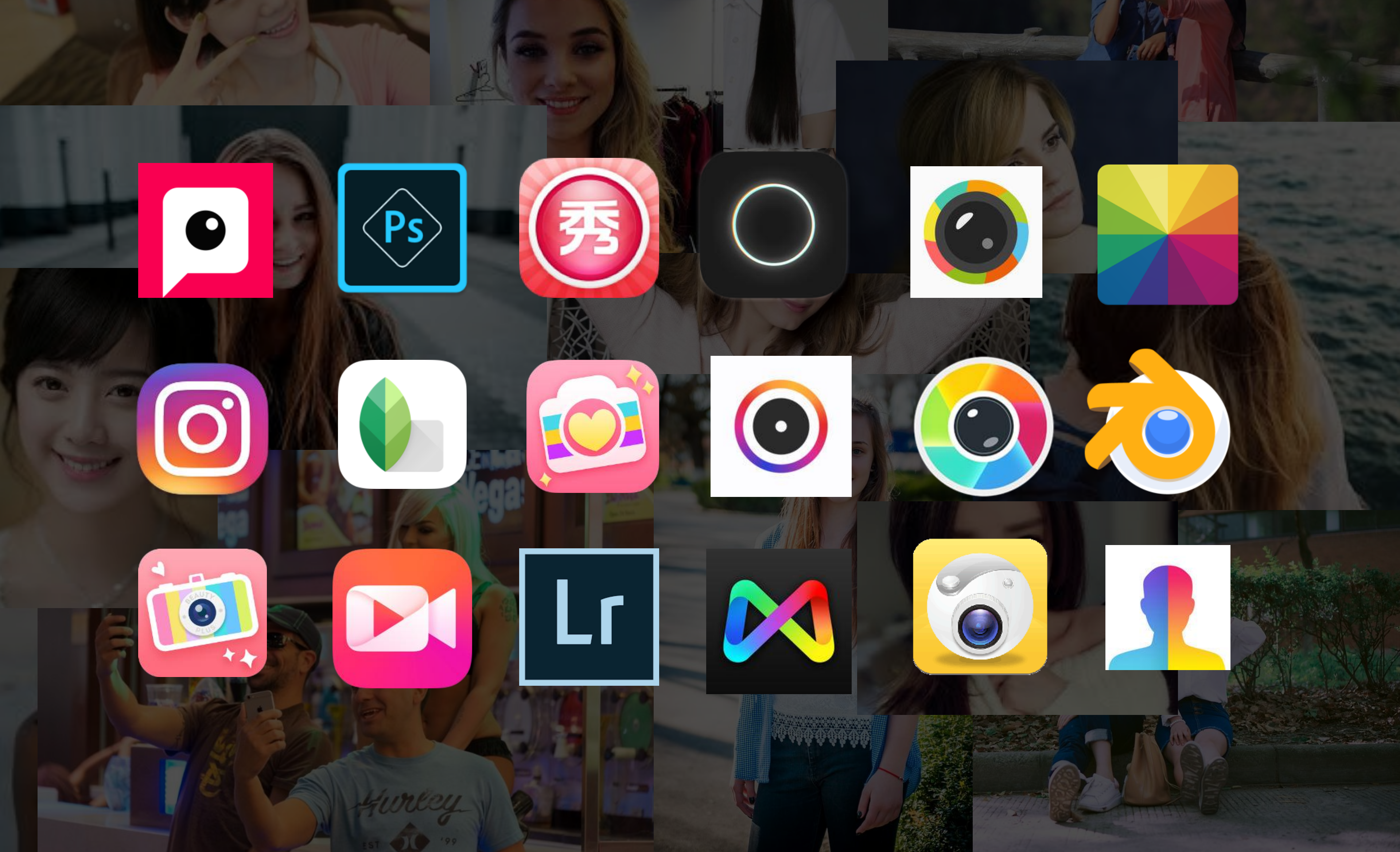
人工智能　腾讯科技　2017-10-18 16:04　收藏　0 评论　分享

腾讯科技讯 被誉为计算机视觉领域三大顶级会议之一的ICCV（另外两个为CVPR、ECCV）近日揭晓收录论文名单，腾讯优图共有12篇论文入选，居业界实验室前列，其中3篇被选做口头报告（Oral），该类论文仅占总投稿数的2.1%（45/2143）。

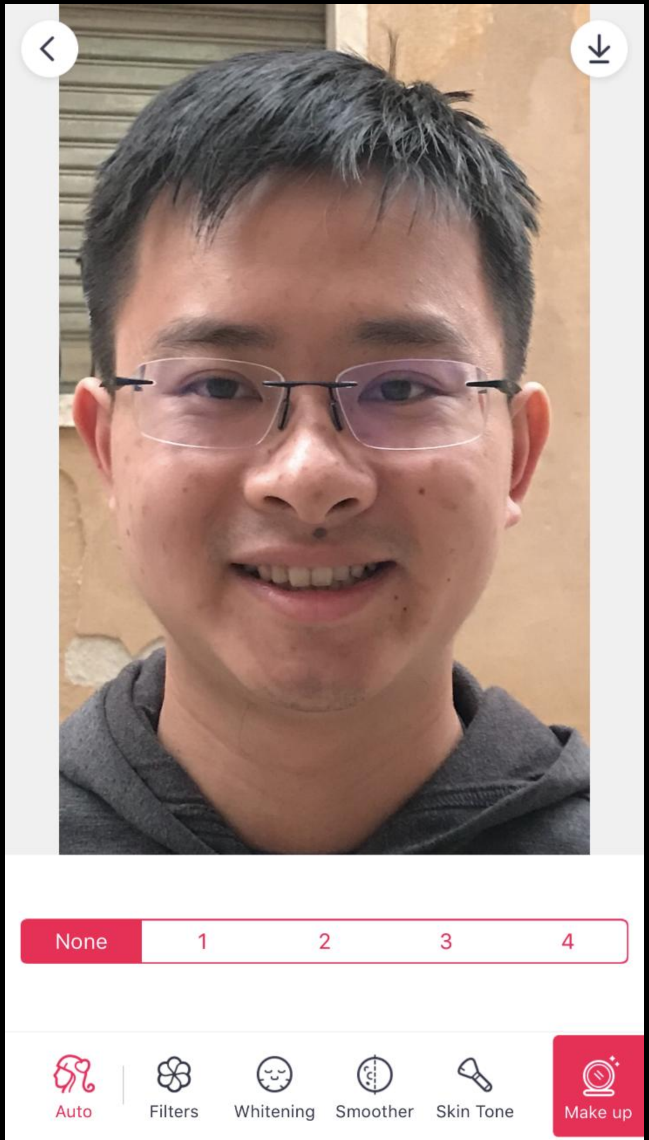

本届 ICCV 共收到2143篇论文投稿，其中621篇被选为大会论文，录用比例29%。其中有45篇口头报告（Oral）和56篇亮点报告（Spotlight）。今年参会人数预计将超过3000人，可见其火爆程度。

# Makeup-Go: Blind Reversion of Portrait Edit

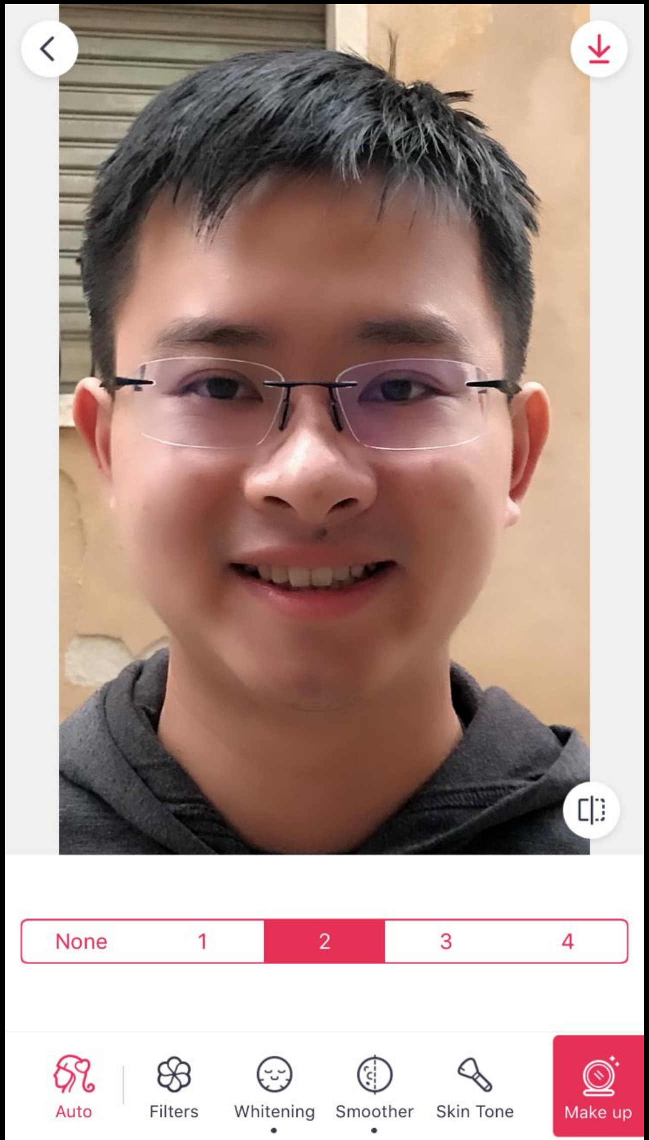
# Popular Digital Editing Tools



# A Digital Edit Example

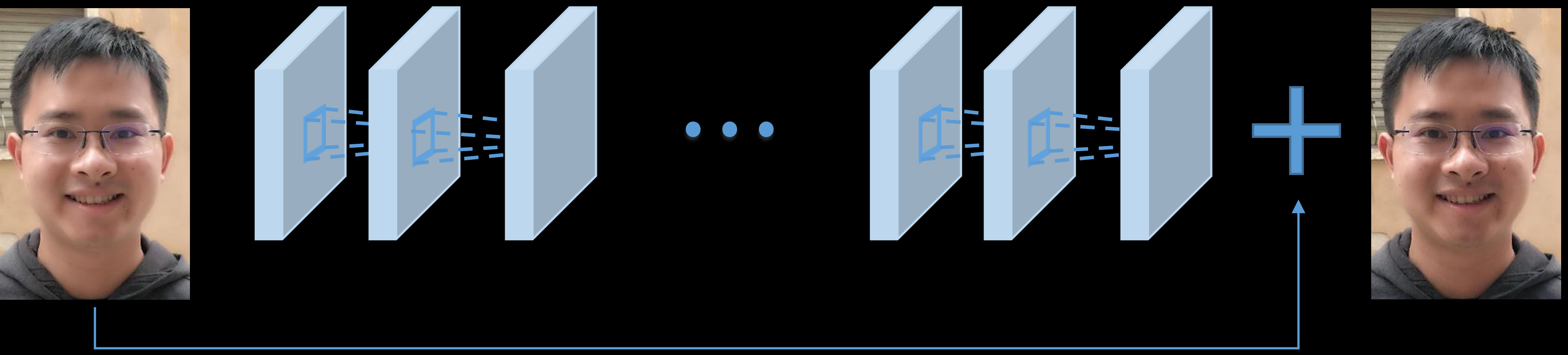


# A Digital Edit Example



What is the difficulty of this task?

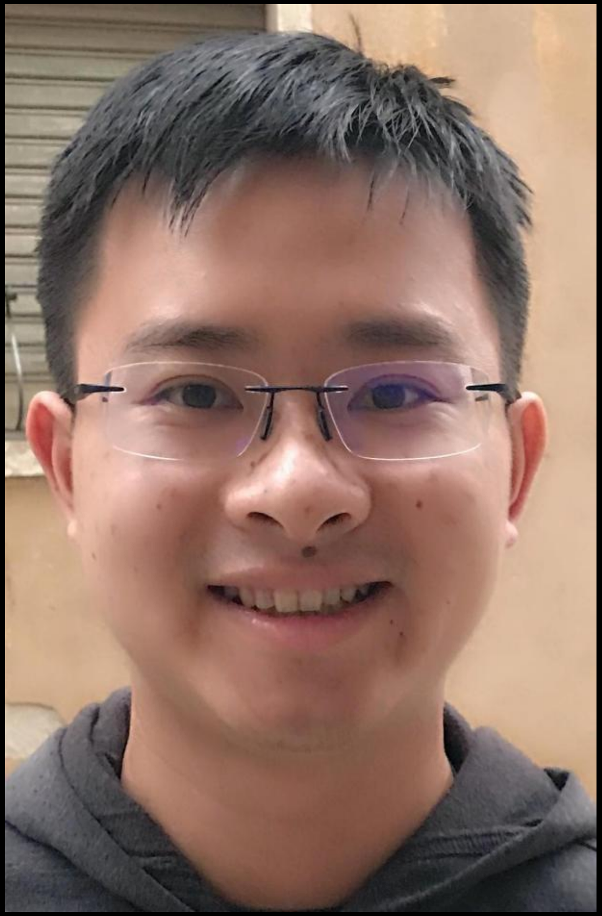
# Existing CNN Cannot Remove Make-up



Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks. " *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

# Result of Applying CNN to Remove Make-up

Input



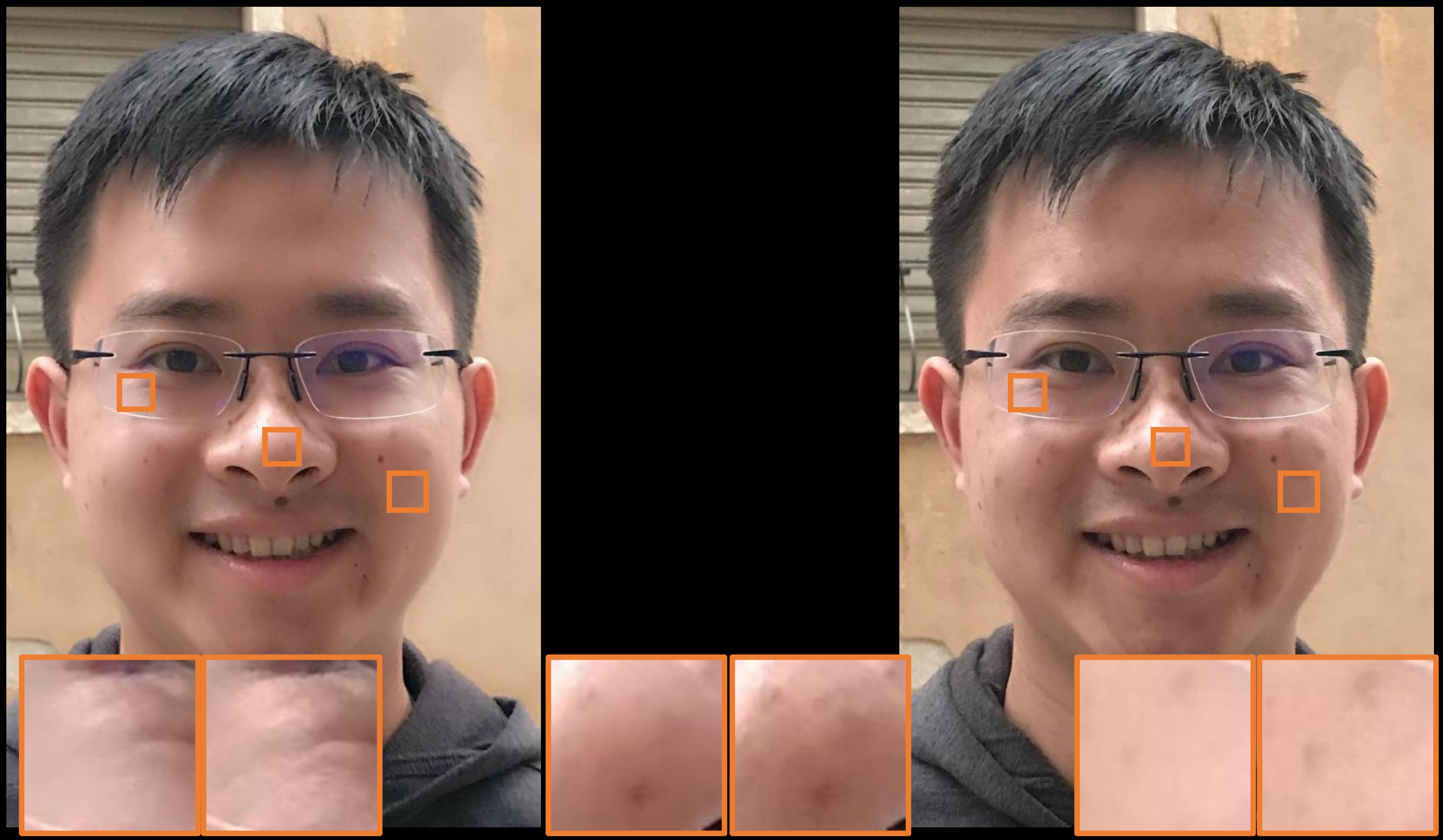
Output

# Result of Applying CNN to Remove Make-up

Ground Truth

Output

# Result of Applying CNN to Remove Make-up



Output

Ground Truth

Why cannot existing CNN achieve this goal?

**Component Domination Effect**

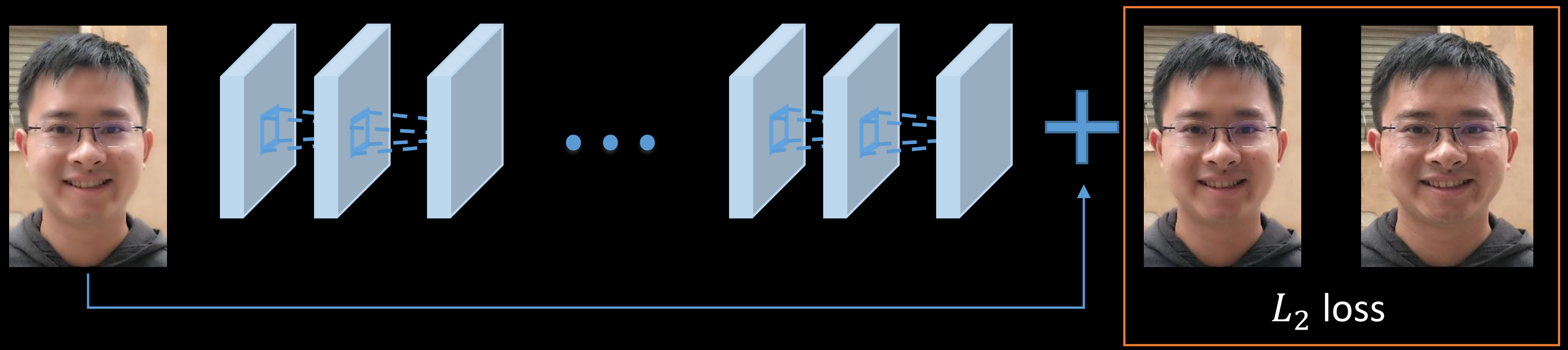
# Analysis of the Loss

$$L = \left|\left|F(x) - y\right|\right|_2^2$$

$x$: input image patch

$F(x)$: network output (vectorized)

$y$: ground truth patch (vectorized)

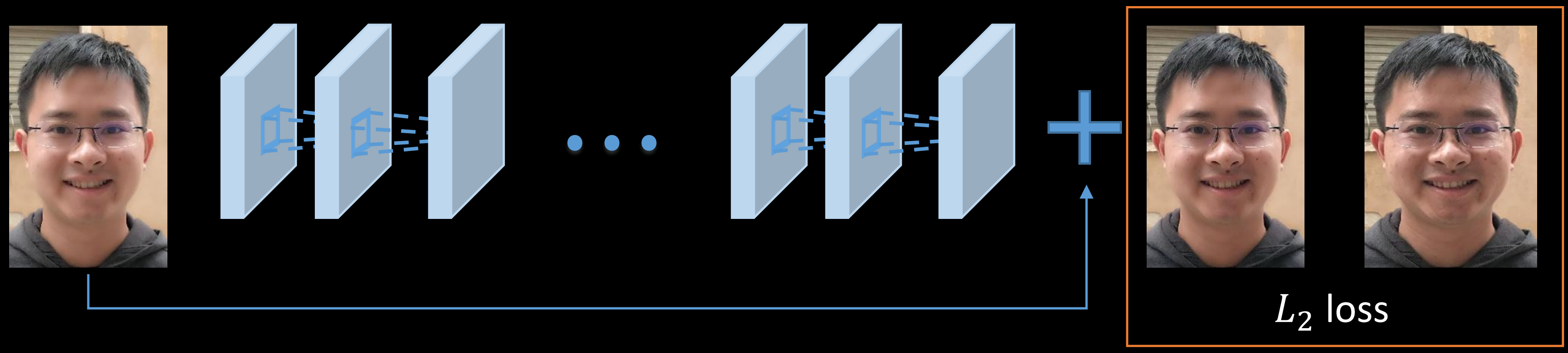


# PCA Components of the Loss

$$L = ||U^T F(x) - U^T y||_2^2$$

$U$: PCA matrix

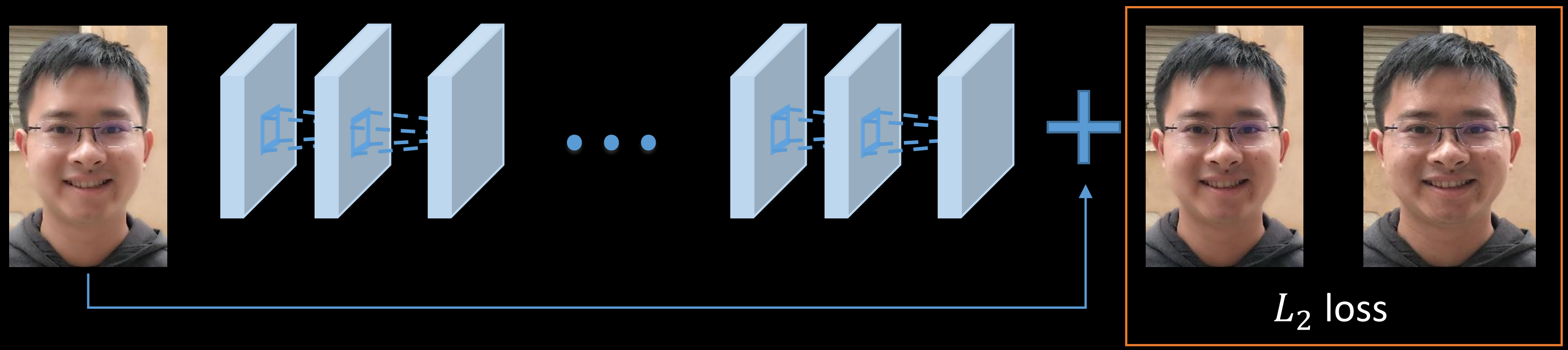
$UU^T = I$



# PCA Components of the Loss

$$L = \left|\left|f_1(x) - u_1^T y\right|\right|_2^2 + \left|\left|f_2(x) - u_2^T y\right|\right|_2^2 + \cdots + \left|\left|f_d(x) - u_d^T y\right|\right|_2^2$$

$$f_i(x) = u_i^T F(x), i = 1,2 \cdots, d$$



# PCA Components of the Loss

$$L = \left|\left|f_1(x) - u_1^T y\right|\right|_2^2 + \left|\left|f_2(x) - u_2^T y\right|\right|_2^2 + \cdots + \left|\left|f_d(x) - u_d^T y\right|\right|_2^2$$

$$f_i(x) = u_i^T F(x), i = 1,2 \cdots, d$$

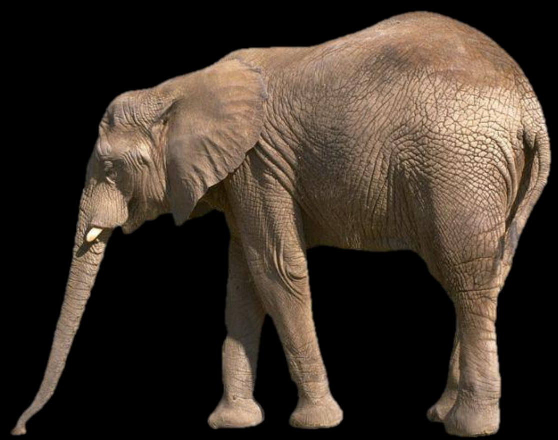
$u_1^T y$

# PCA Components of the Loss

$$L = \left|\left|f_1(x) - u_1^T y\right|\right|_2^2 + \left|\left|f_2(x) - u_2^T y\right|\right|_2^2 + \cdots + \left|\left|f_d(x) - u_d^T y\right|\right|_2^2$$

$$f_i(x) = u_i^T F(x), i = 1,2 \cdots, d$$

$u_1^T y$



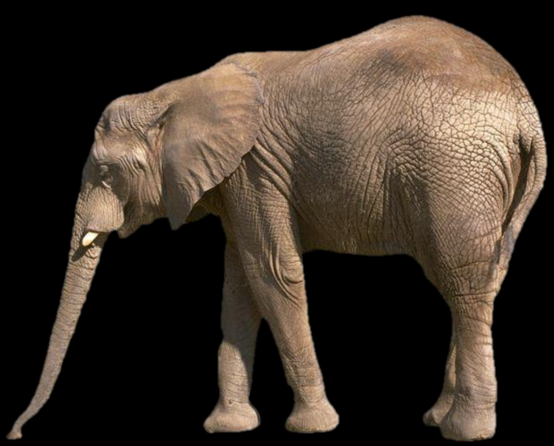
$u_2^T y$

# PCA Components of the Loss

$$L = \left\|f_1(x) - u_1^T y\right\|_2^2 + \left\|f_2(x) - u_2^T y\right\|_2^2 + \cdots + \left\|f_d(x) - u_d^T y\right\|_2^2$$

$$f_i(x) = u_i^T F(x), i = 1,2 \cdots, d$$

$u_1^T y$



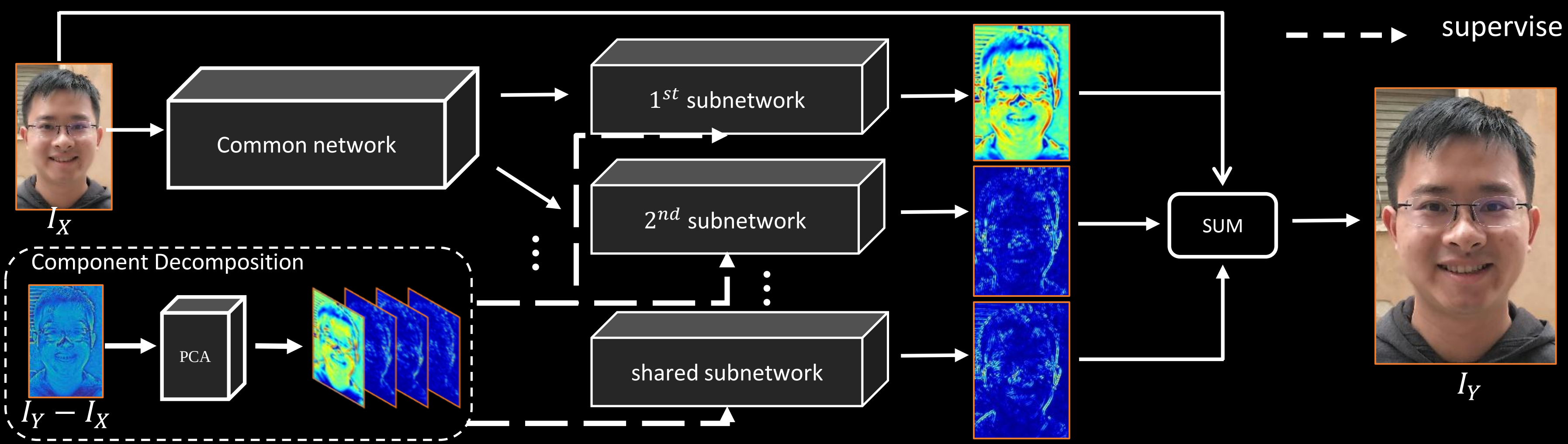
$u_2^T y$

…

$u_d^T y$

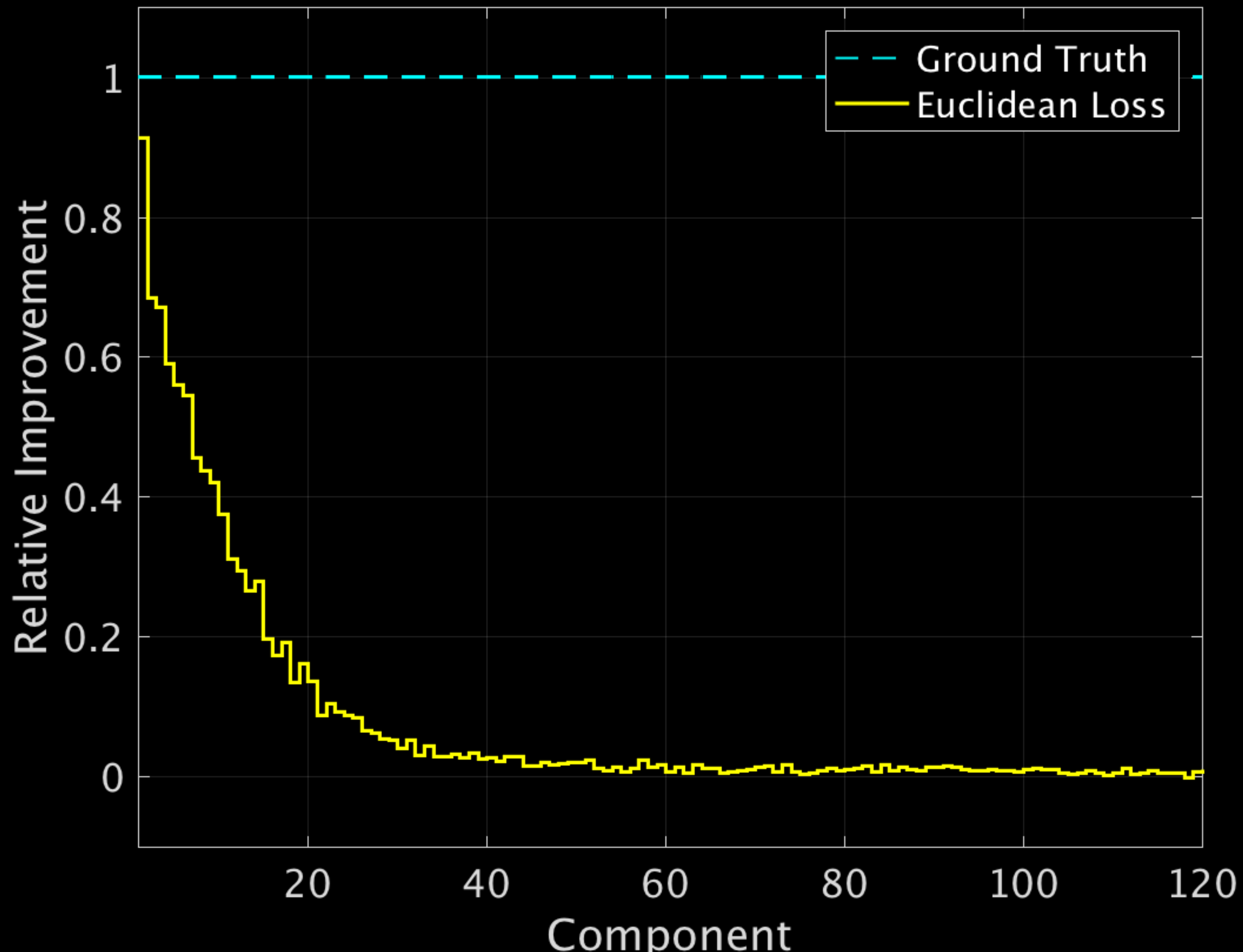
# Our Framework

# Component Regression Network

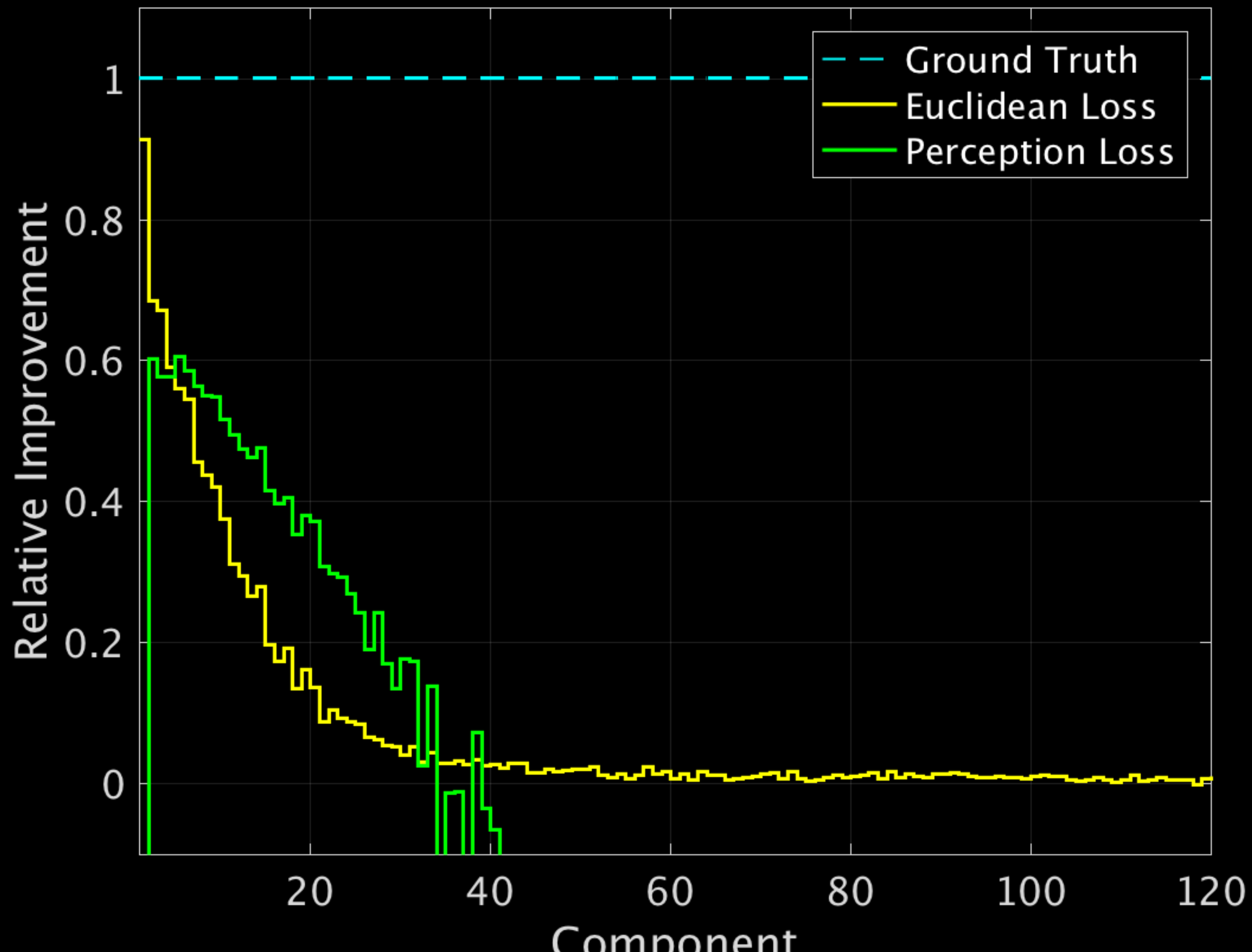


# Experimental Results

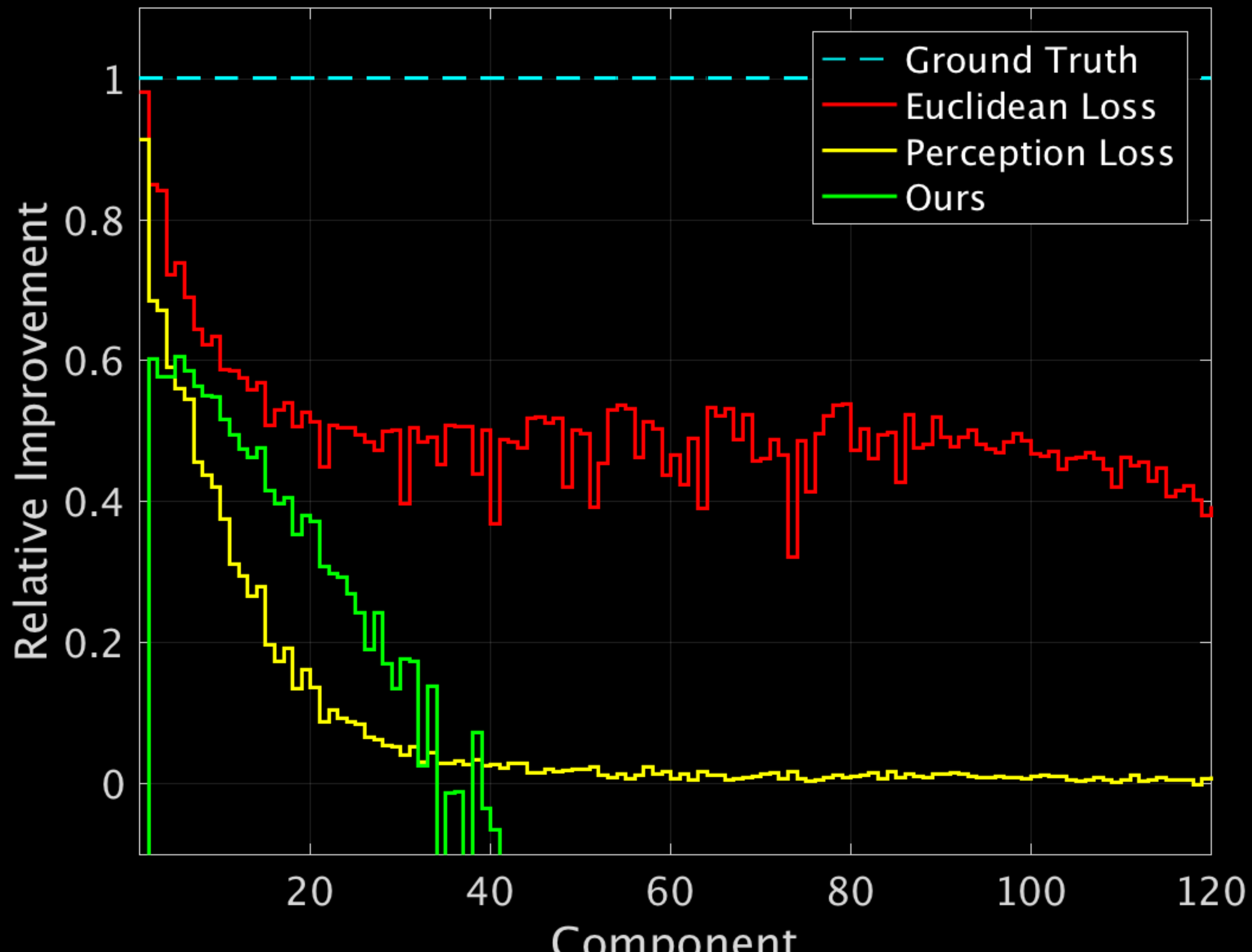
# Component-level Comparison

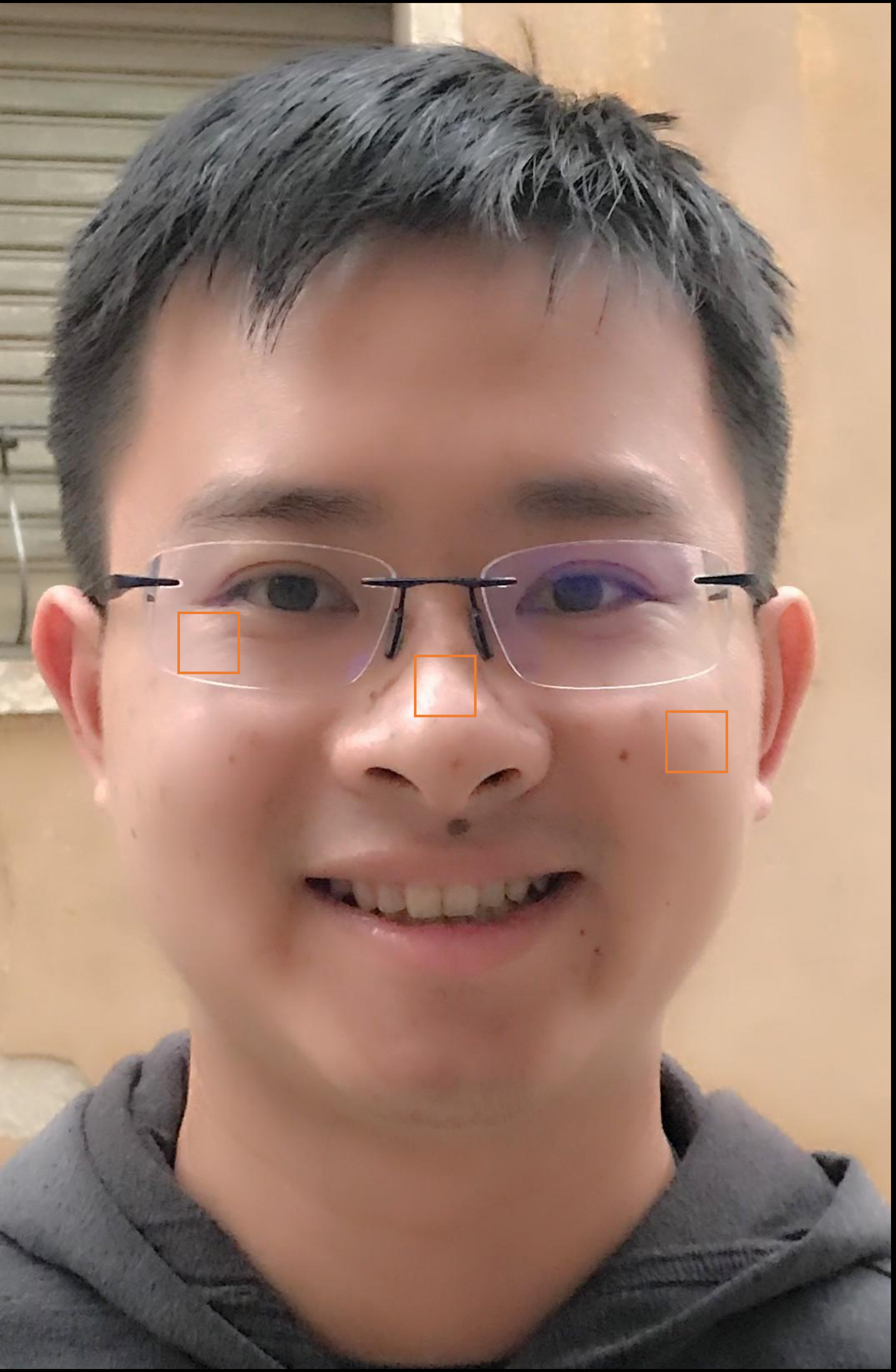


# Component-level Comparison



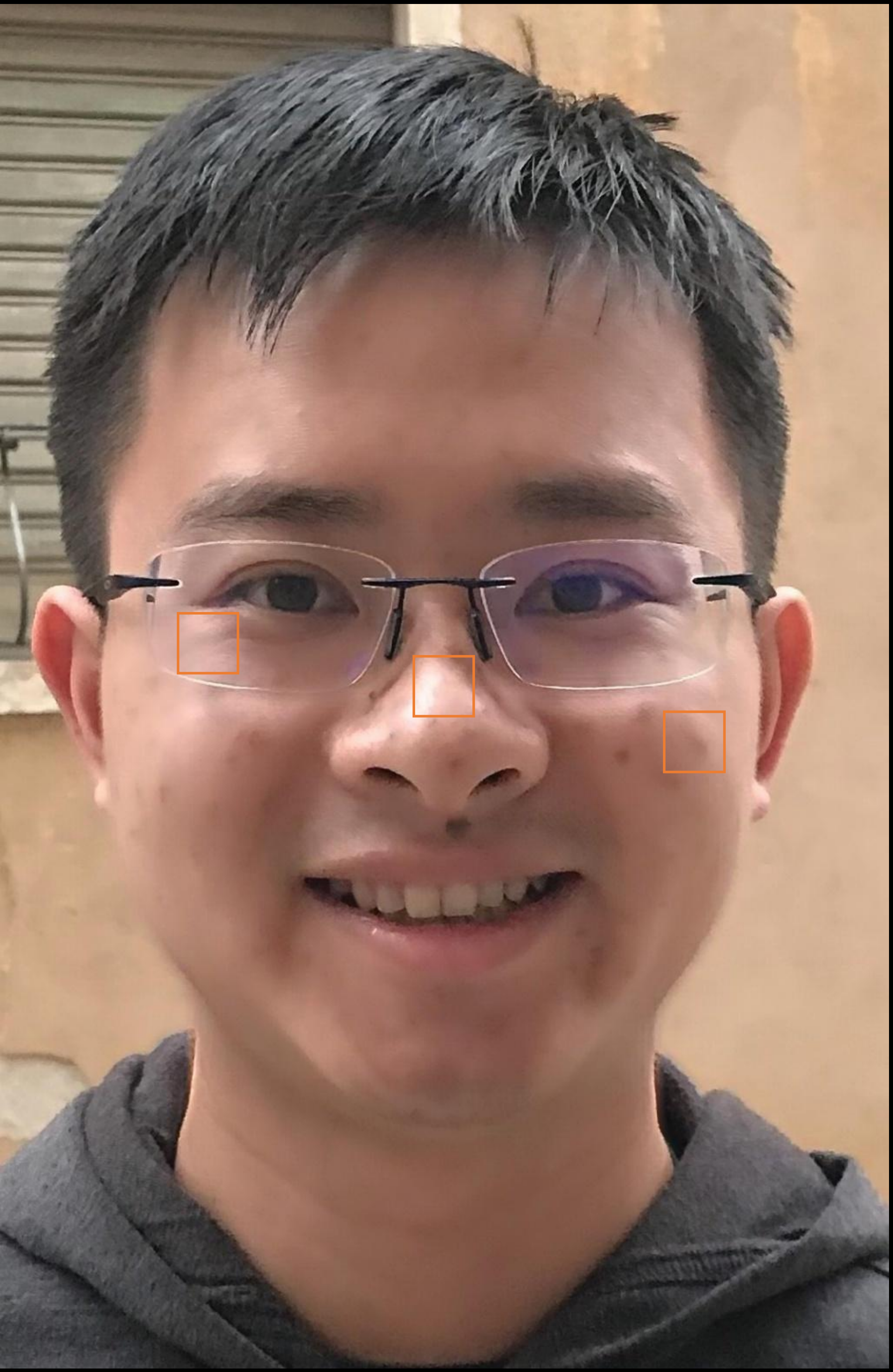
# Component-level Comparison



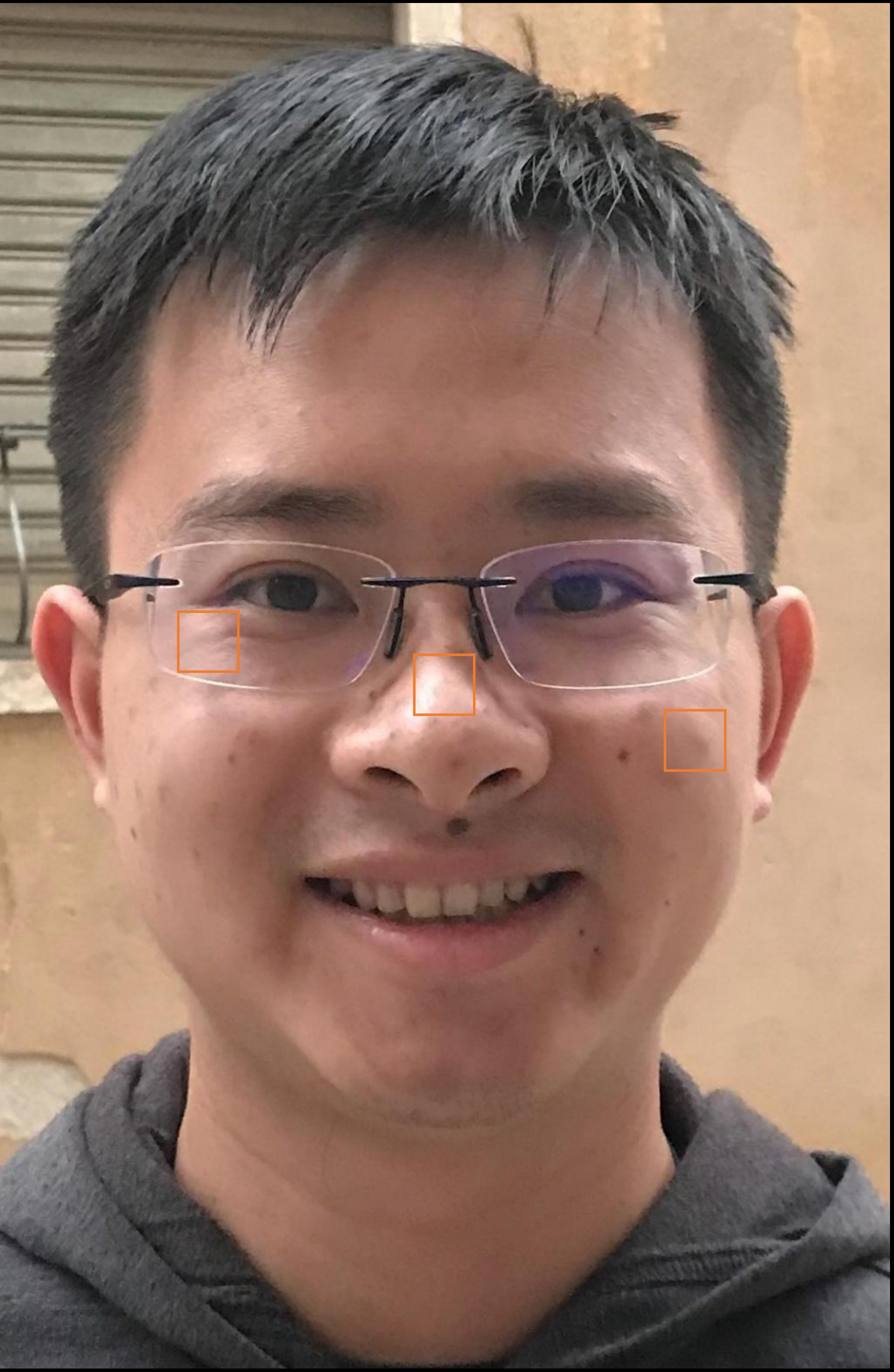


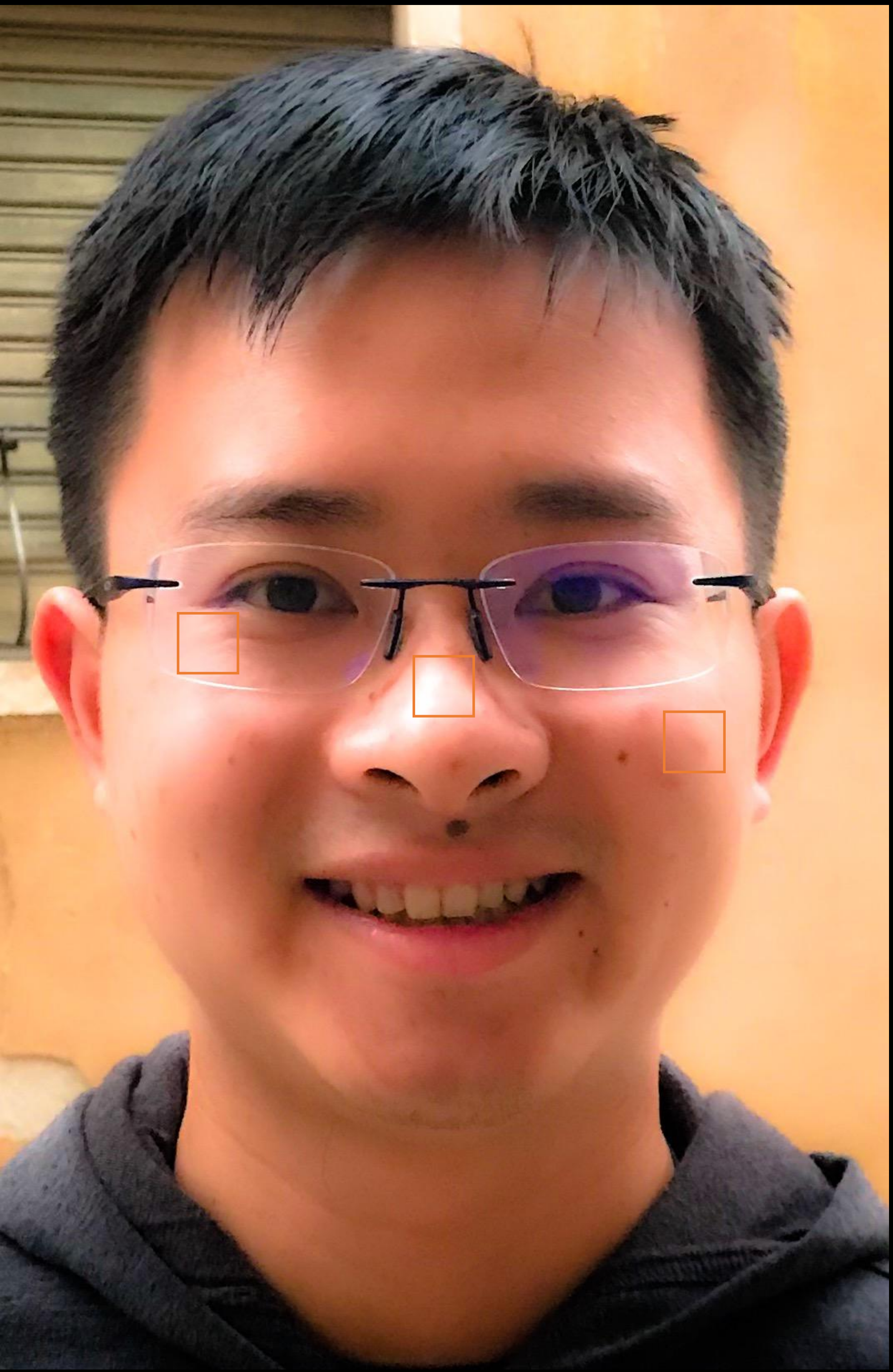
# Input Touched Image

# Our Result

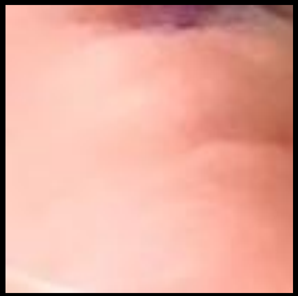
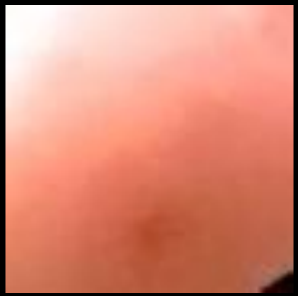


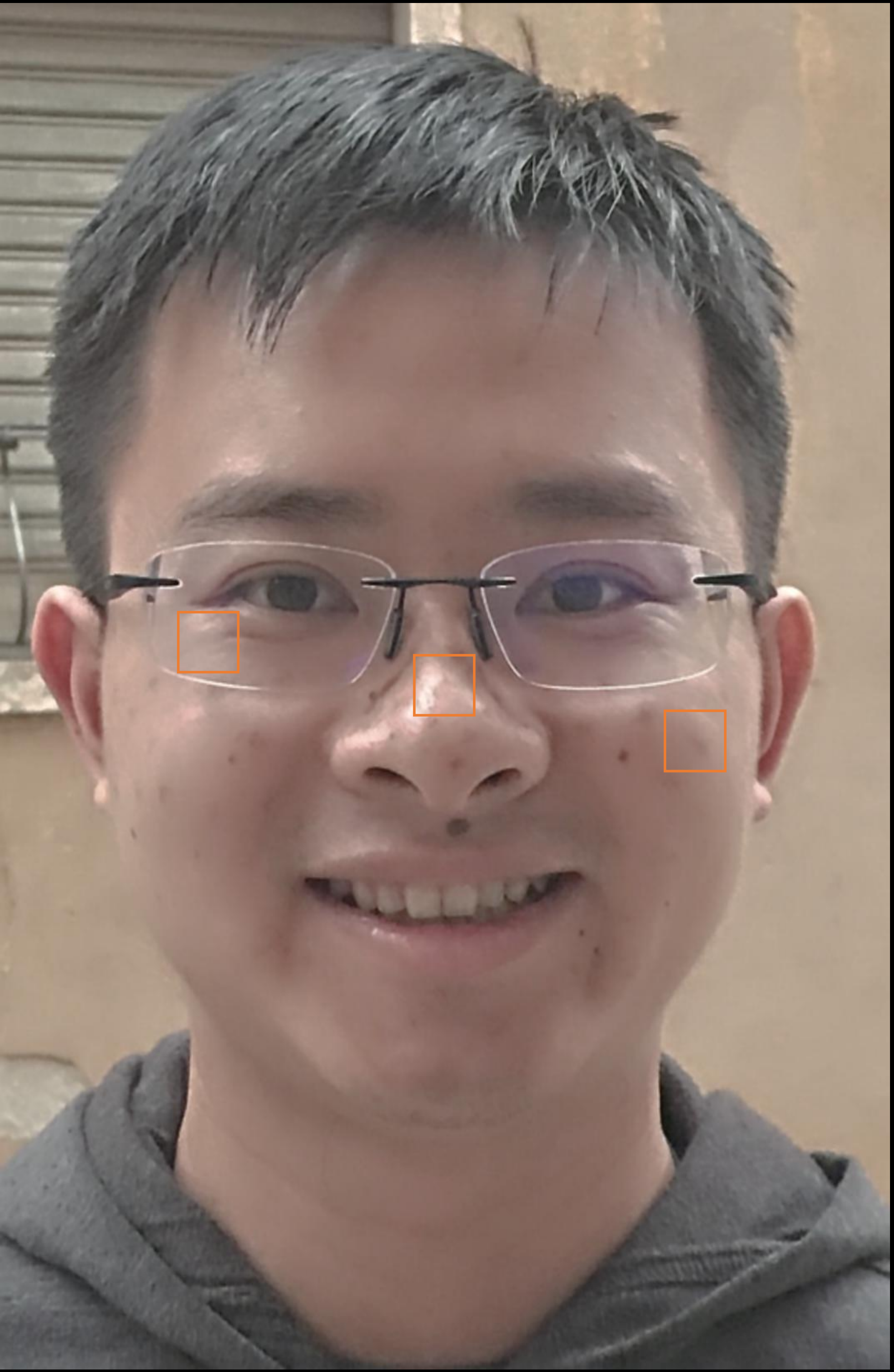
# Ground Truth



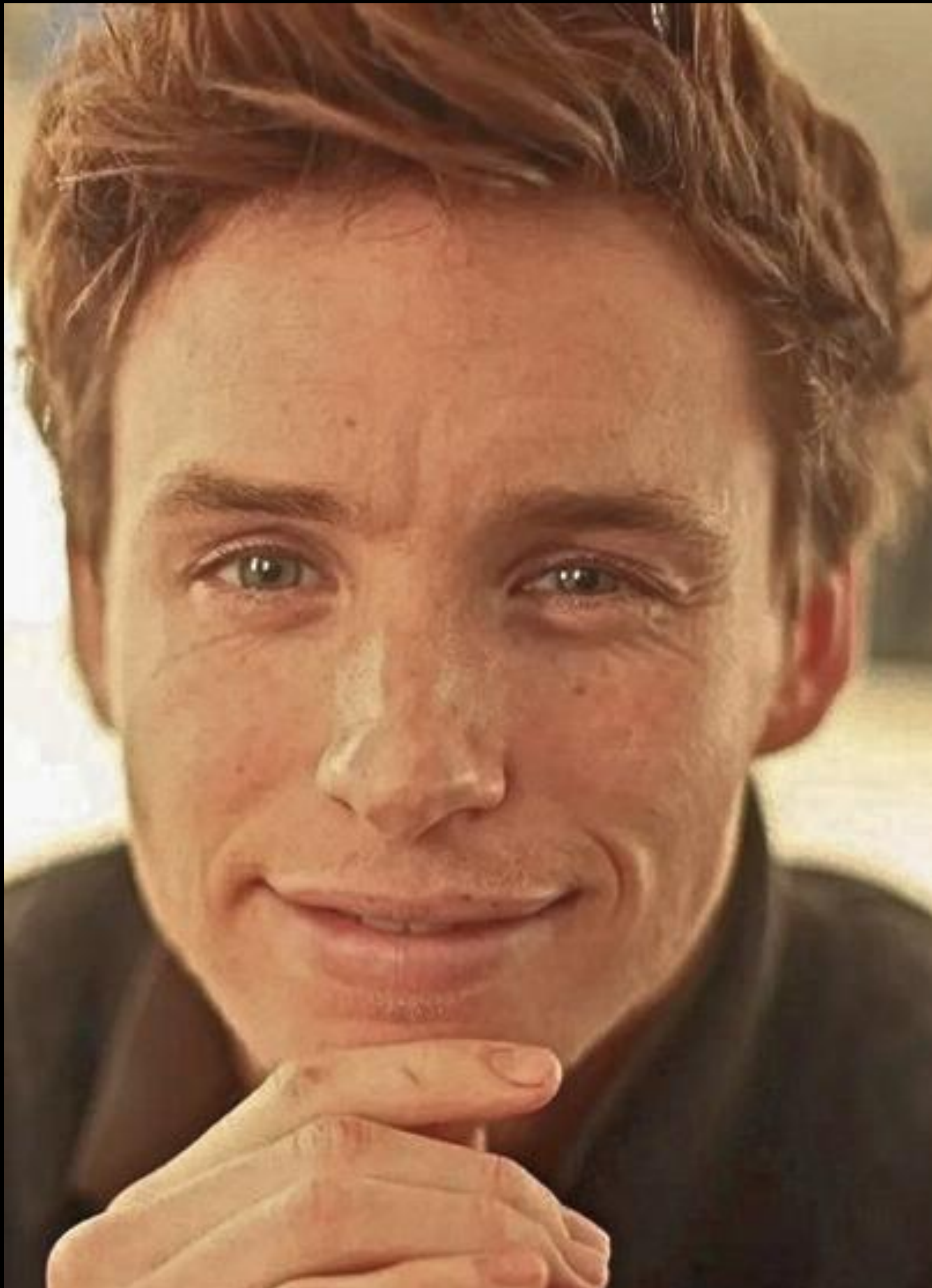


# Overly Touched

# Our result





# Not-That-Good Result

Ground Truth

Touched

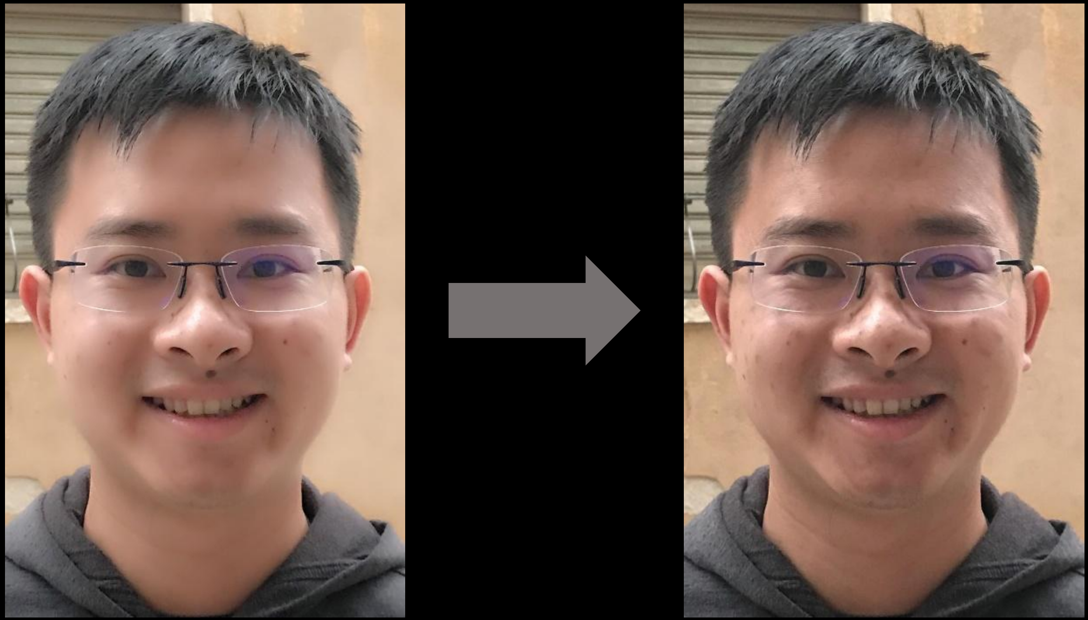
# Not-That-Good Result

Ground Truth

Output

# Summary

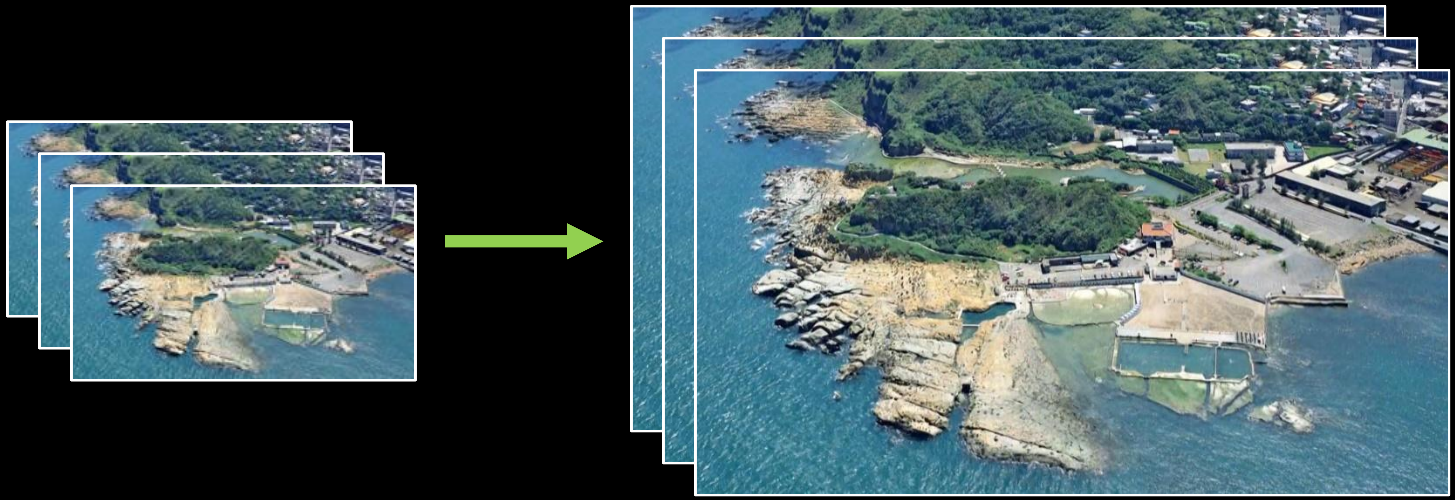
- We discovered the component domination effect.
- We proposed a Component Regression Network to tackle the problem.



# Detail-revealing Deep Video Super-resolution

# Motivation

- Old and Fundamental
  - Several decades ago [Huang et al, 1984] → near recent
- Many Applications
  - HD video generation from low-res sources

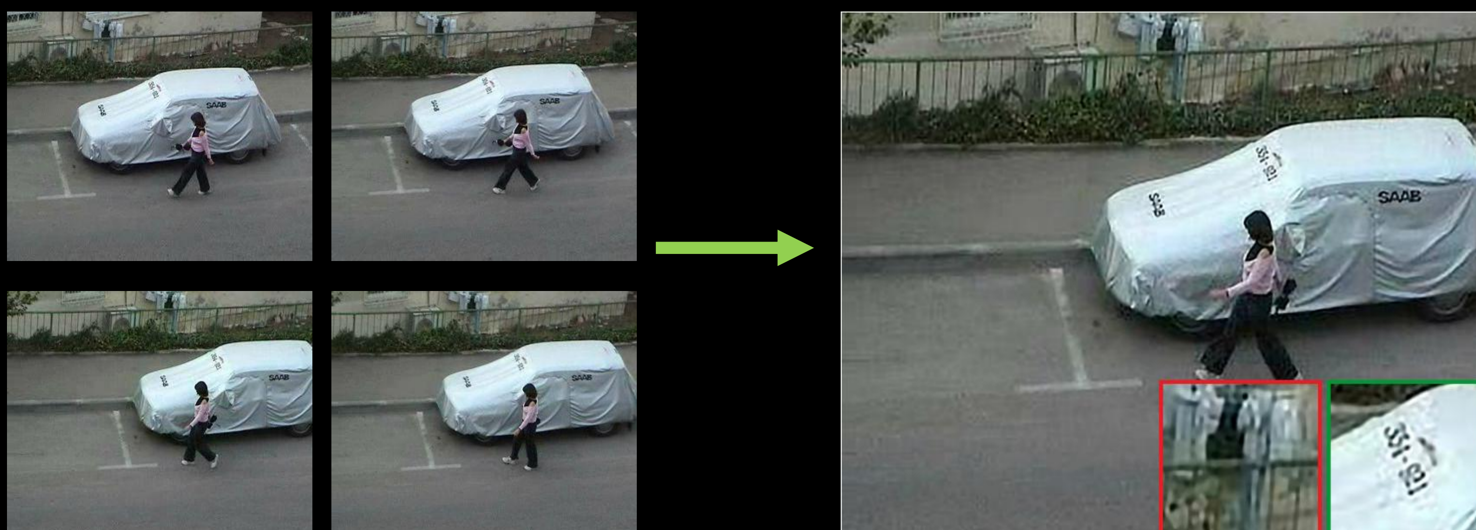


# Motivation

- Old and Fundamental
  - Several decades ago [Huang et al, 1984] → near recent
- Many Applications
  - HD video generation from low-res sources
  - Video enhancement with details

# Motivation

- Old and Fundamental
  - Several decades ago [Huang et al, 1984] → near recent
- Many Applications
  - HD video generation from low-res sources
  - Video enhancement with details
  - Text/object recognition in surveillance videos



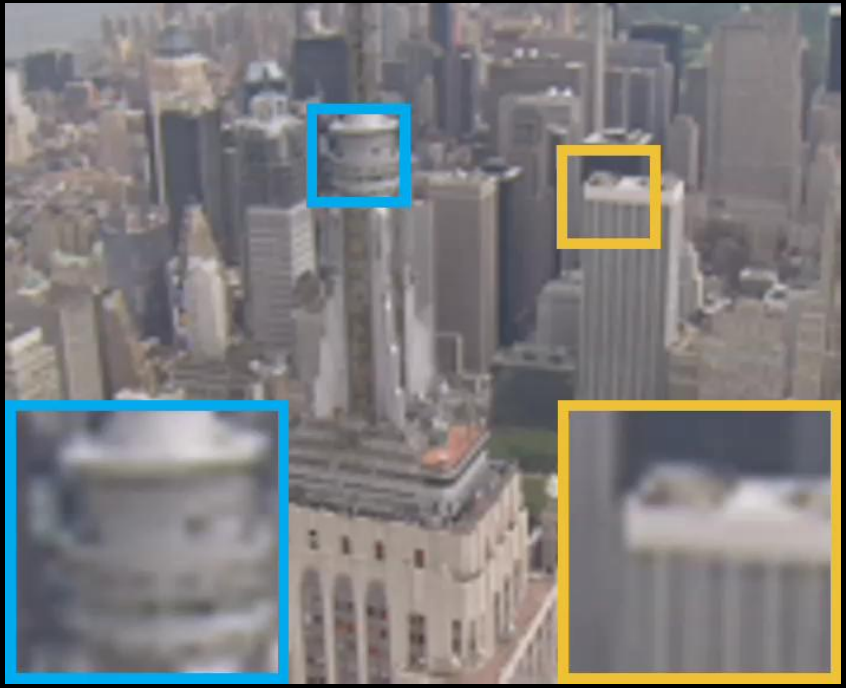
# Previous Work

- Image SR
  - Traditional: [Freeman et al, 2002], [Glasner et al, 2009], [Yang et al, 2010], etc.
  - CNN-based: SRCNN [Dong et al, 2014], VDSR [Kim et al, 2016], FSRCNN [Dong et al, 2016], etc.

- Video SR
  - Traditional: 3DSKR [Takeda et al, 2009], BayesSR [Liu et al, 2011], MFSR [Ma et al, 2015], etc.
  - CNN-based: DESR [Liao et al, 2015], VSRNet [Kappeler, et al, 2016], [Caballero et al, 2016], etc.

# Remaining Challenges

- Effectiveness
  - How to make good use of multiple frames?



Bicubic x4

Misalignment

Large motion

Occlusion

Data from Vid4 [Ce Liu *et al.*]

# Remaining Challenges

- Effectiveness
  - How to make good use of multiple frames?
  - Are the generated details real?

Bicubic x4

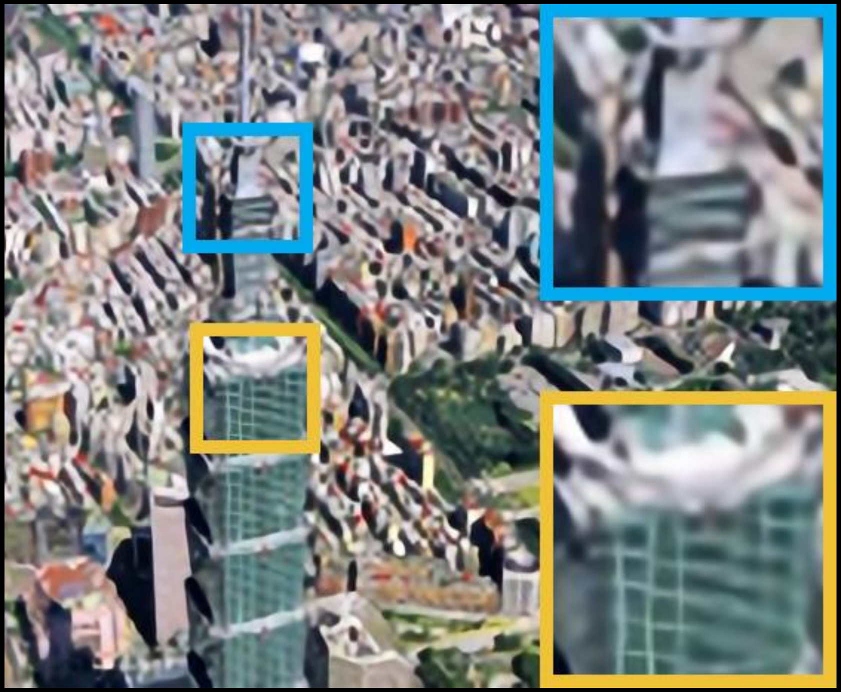
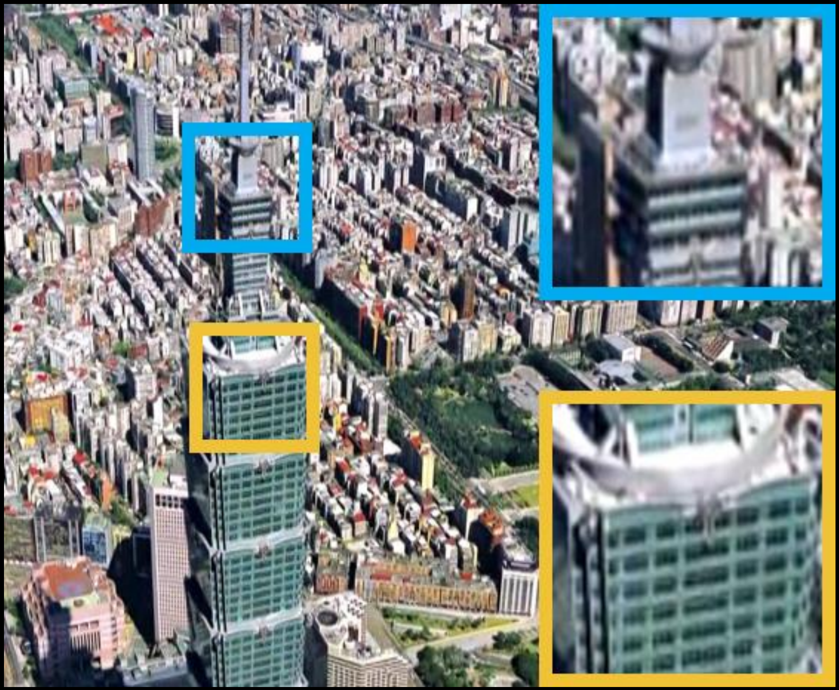


Image SR

# Remaining Challenges

- Effectiveness
  - How to make good use of multiple frames?
  - Are the generated details real?



Truth

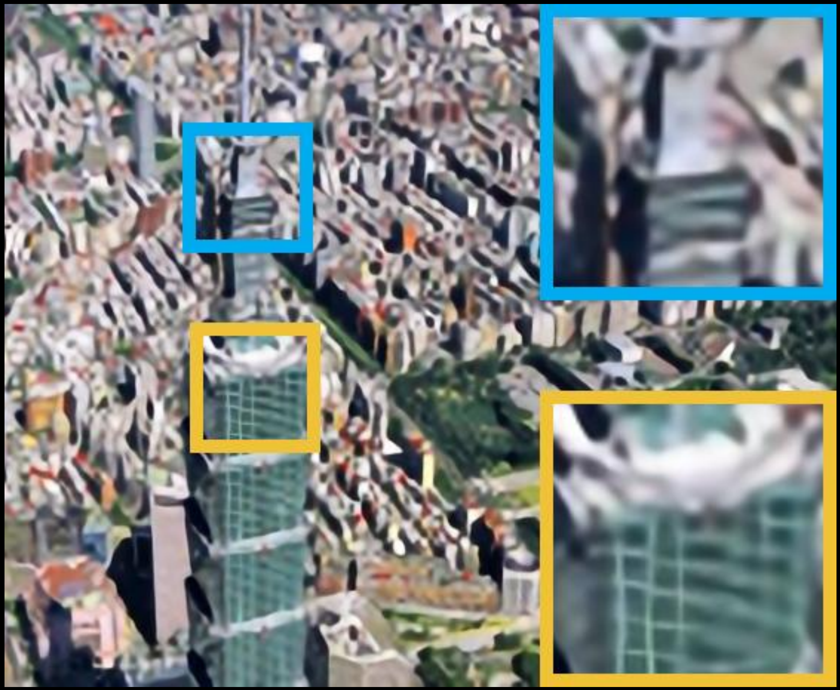
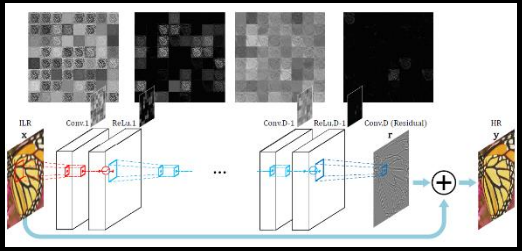


Image SR

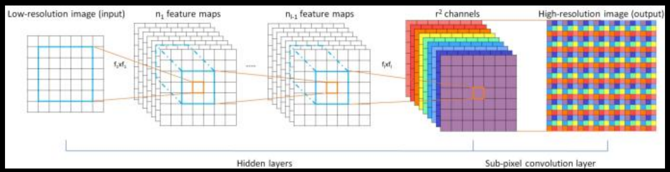
# Remaining Challenges

- Effectiveness
  - How to make good use of multiple frames?
  - Are the generated details real?

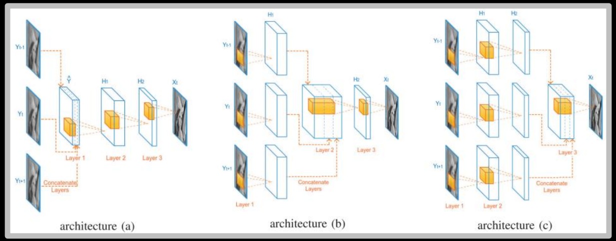
- Model Issues
  - One model for one setting




VDSR [Kim et al., 2016]




ESPCN [Shi et al., 2016]




VSRNet [Kappeler et al, 2016]

# Remaining Challenges

- Effectiveness
  - How to make good use of multiple frames?
  - Are the generated details real?
- Model Issues
  - One model for one setting
  - Intensive parameter tuning
  - Slow

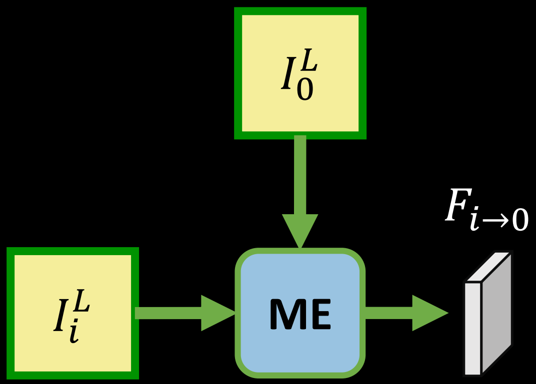
# Our Method

- Advantages
  - Better use of sub-pixel motion
  - Promising results both visually and quantitatively
- Fully Scalable
  - Arbitrary input size
  - Arbitrary scale factor
  - Arbitrary temporal frames


Data from Vid4 [Ce Liu *et al.*]

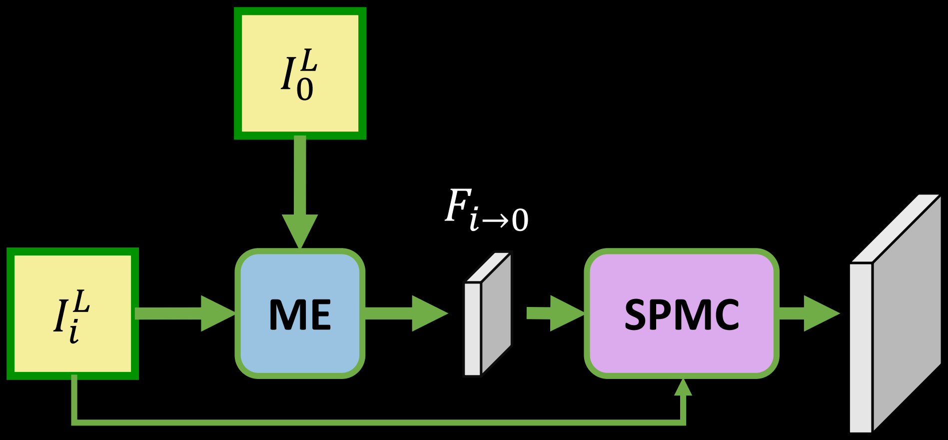
# Our Method

- Motion Estimation



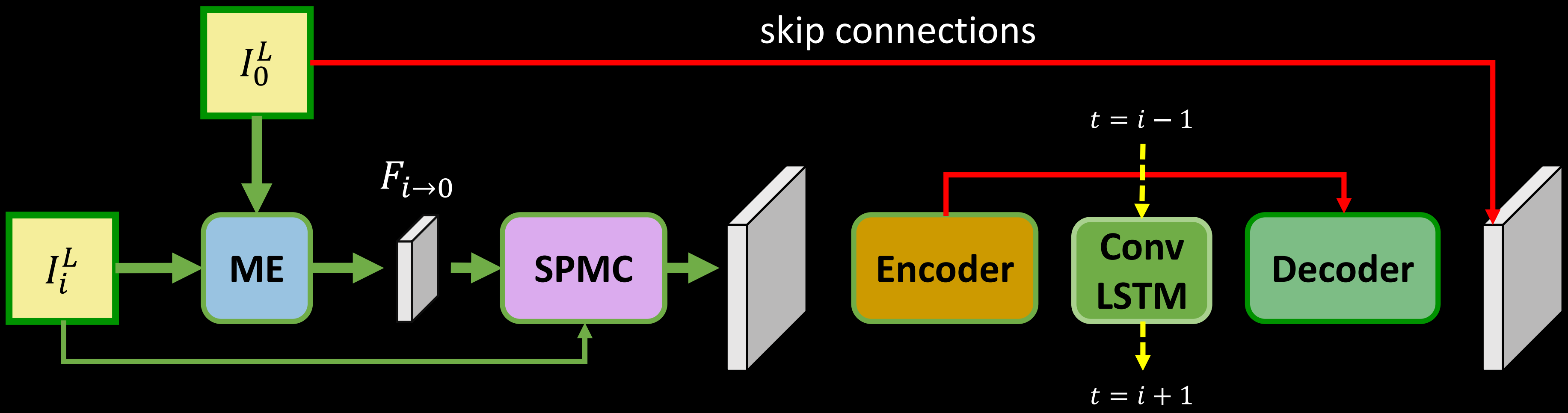
# Our Method

- Sub-pixel Motion Compensation (SPMC) Layer

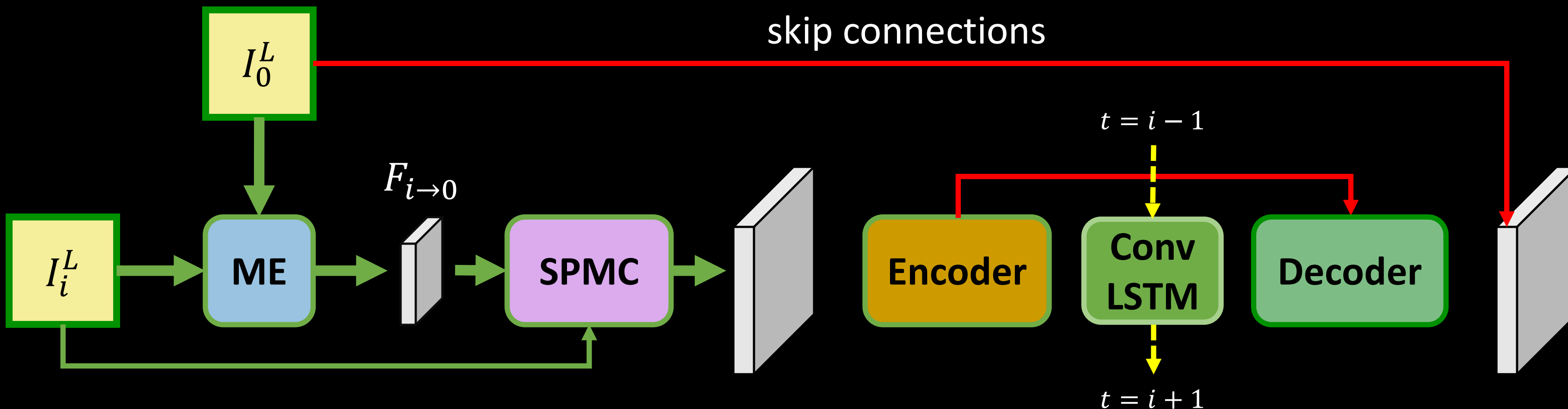


# Our Method

- Detail Fusion Net



# Arbitrary Input Size

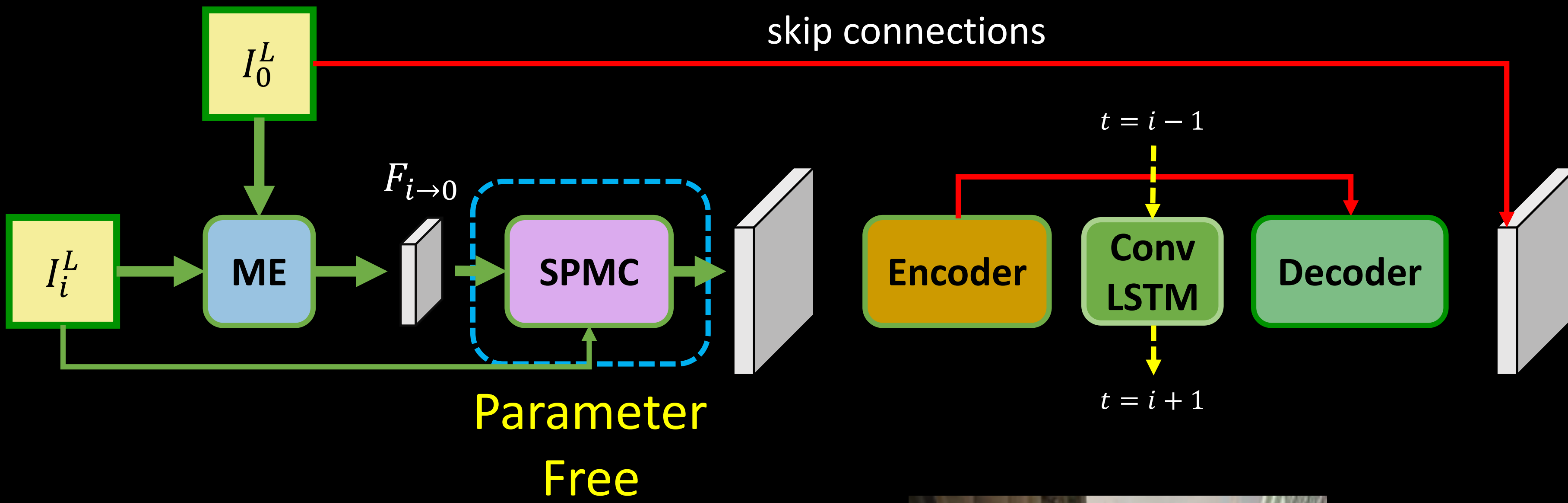



Fully convolutional



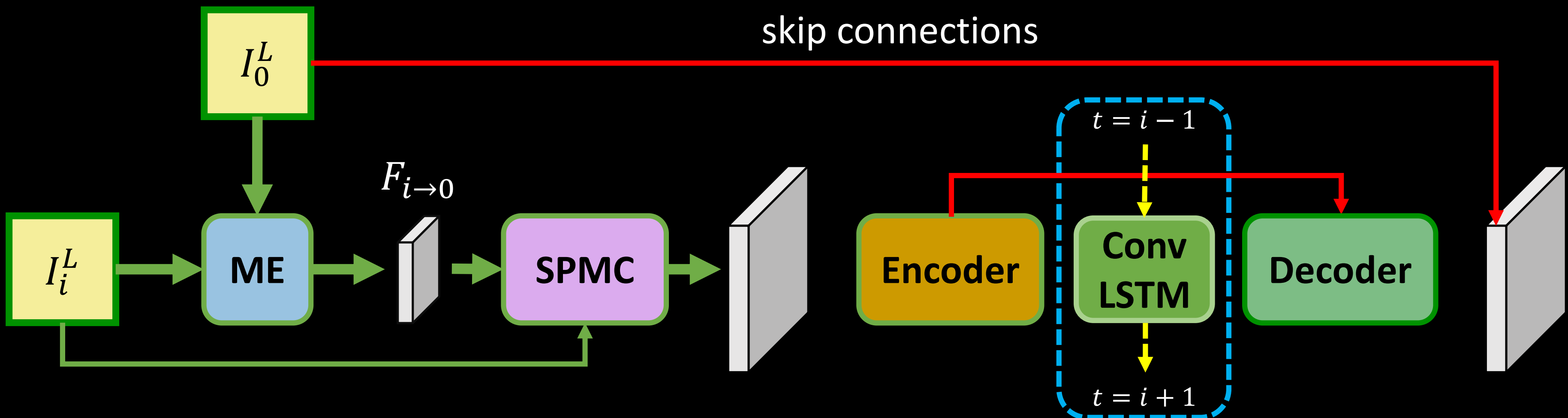
Input size:

# Arbitrary Scale Factors




2× 3× 4×

# Arbitrary Temporal Length




3 frames

5 frames

# Analysis

- Details from multi-frames

Output (identical)

3 identical frames

# Analysis

- Details from multi-frames

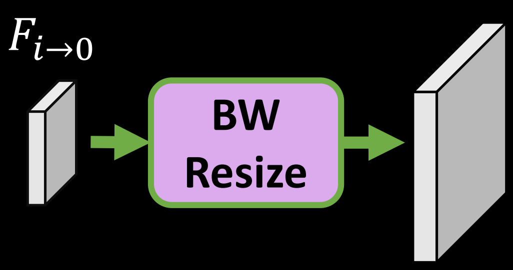
Output (identical)


Output (consecutive)

3 consecutive frames

# Analysis

- Ablation Study: SPMC Layer *v.s.* Baseline

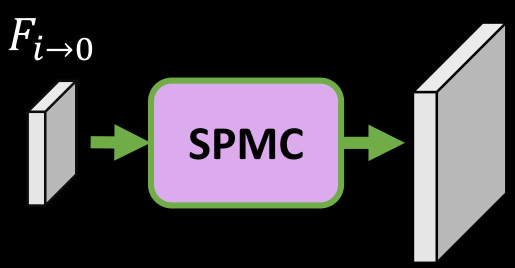



Backward warping
+
Resize
(baseline)

Output (baseline)

# Analysis

- Ablation Study: SPMC Layer *v.s.* Baseline

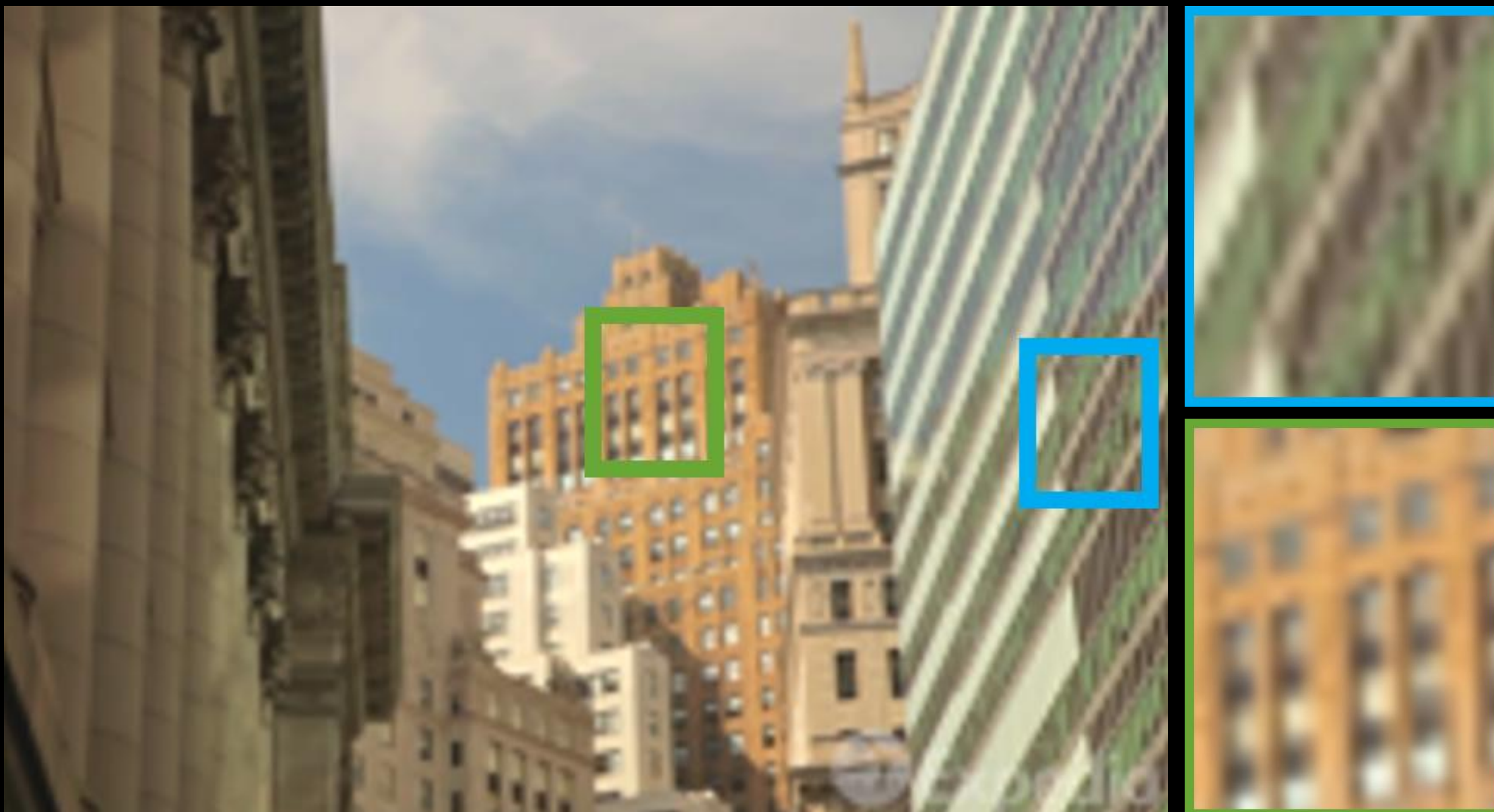



SPMC

Output (baseline)

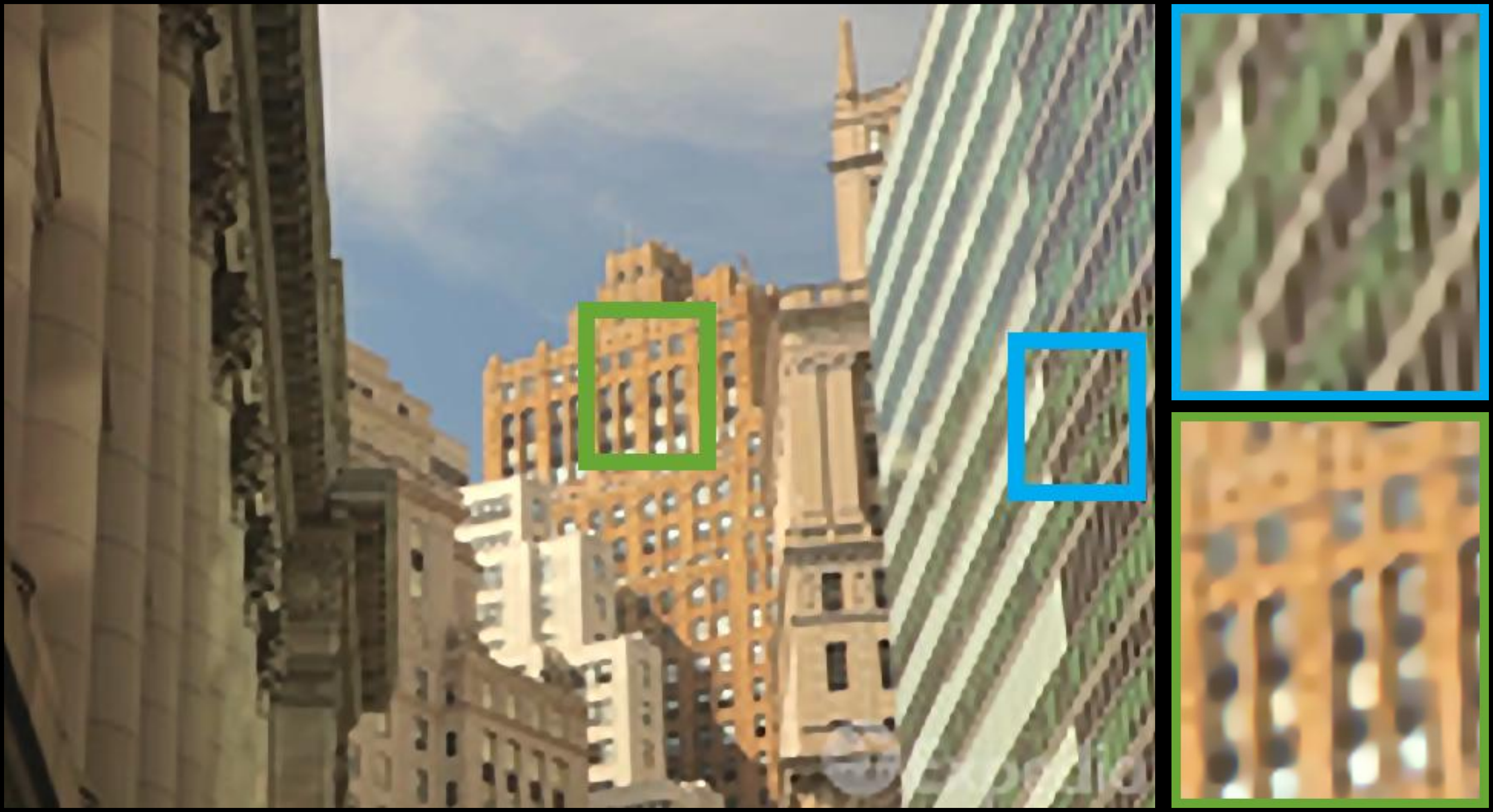
Output (SPMC)

# Comparisons



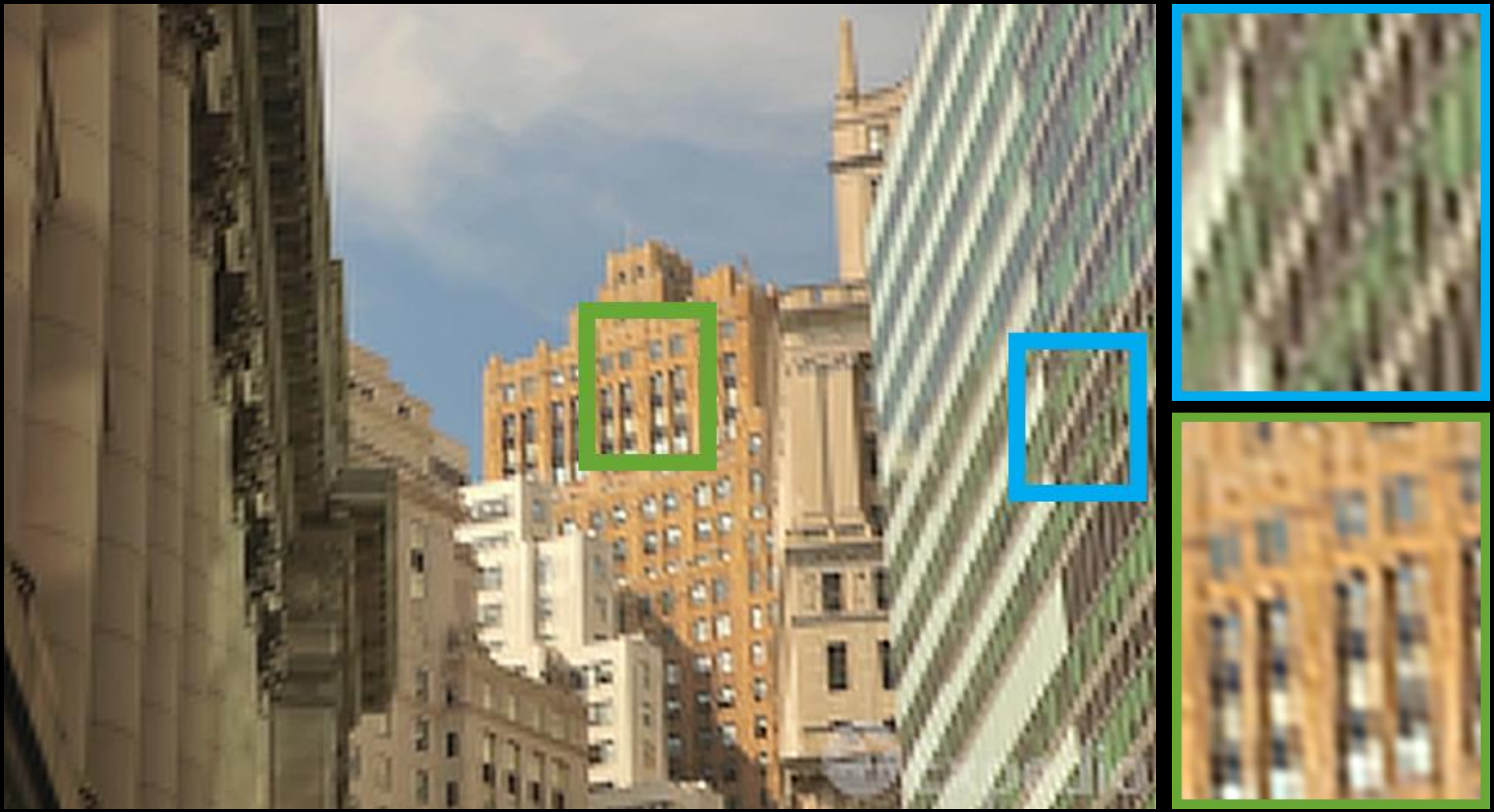
Bicubic x4

# Comparisons



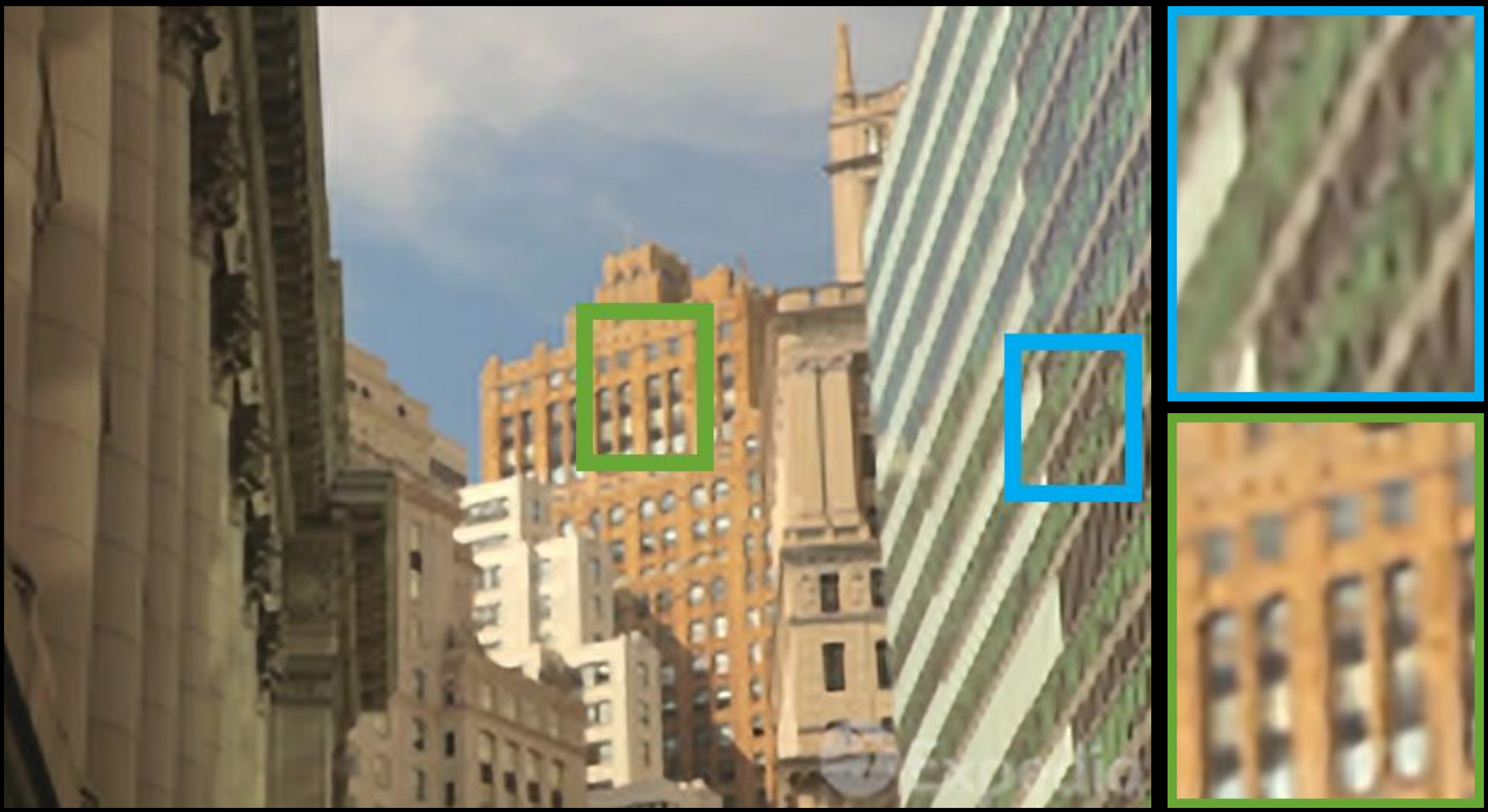
BayesSR [Liu et al, 2011; Ma et al., 2015]

# Comparisons



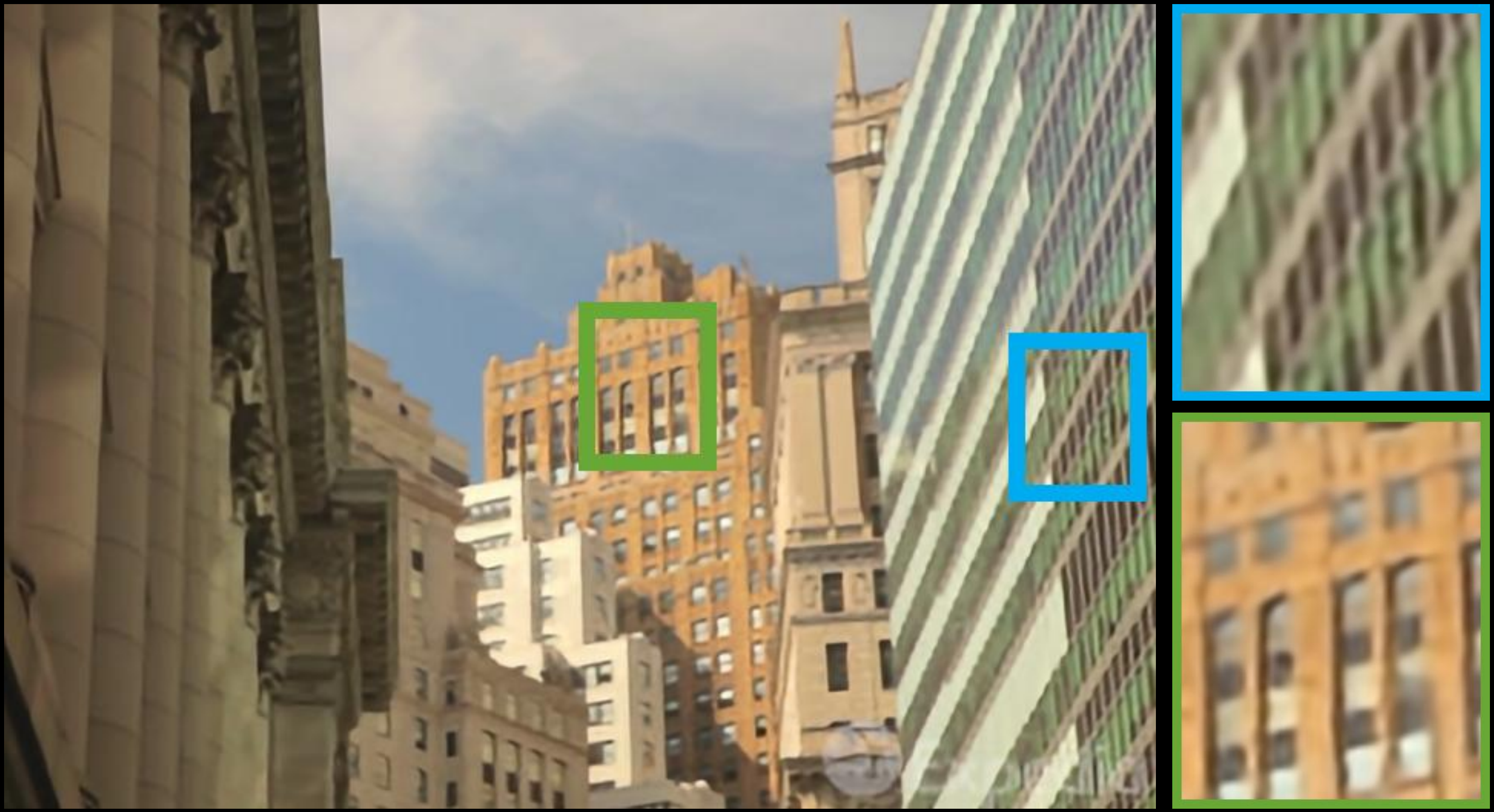
DESR [Liao et al., 2015]

# Comparisons



VSRNet [Kappeler et al, 2016]

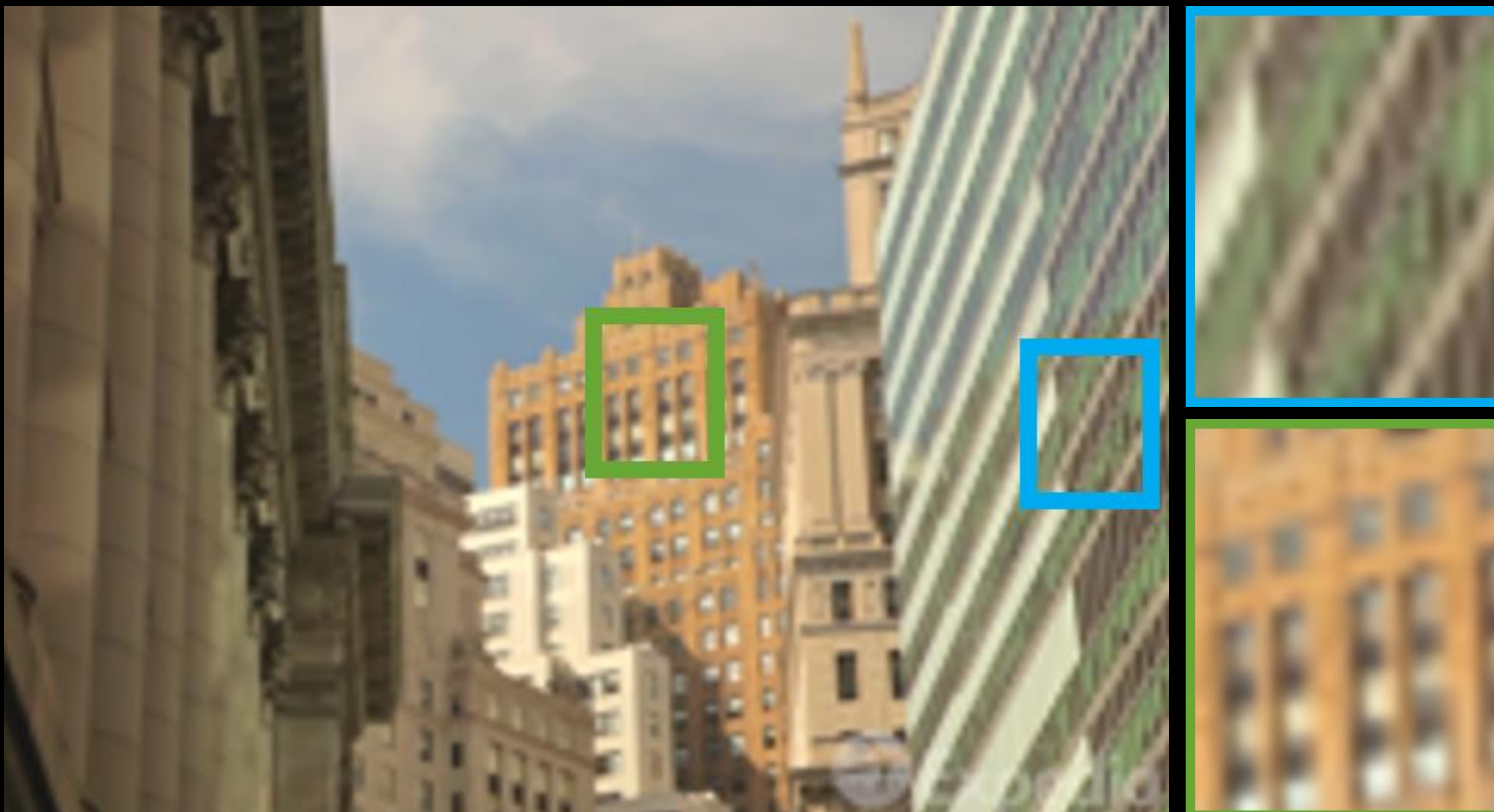
# Comparisons





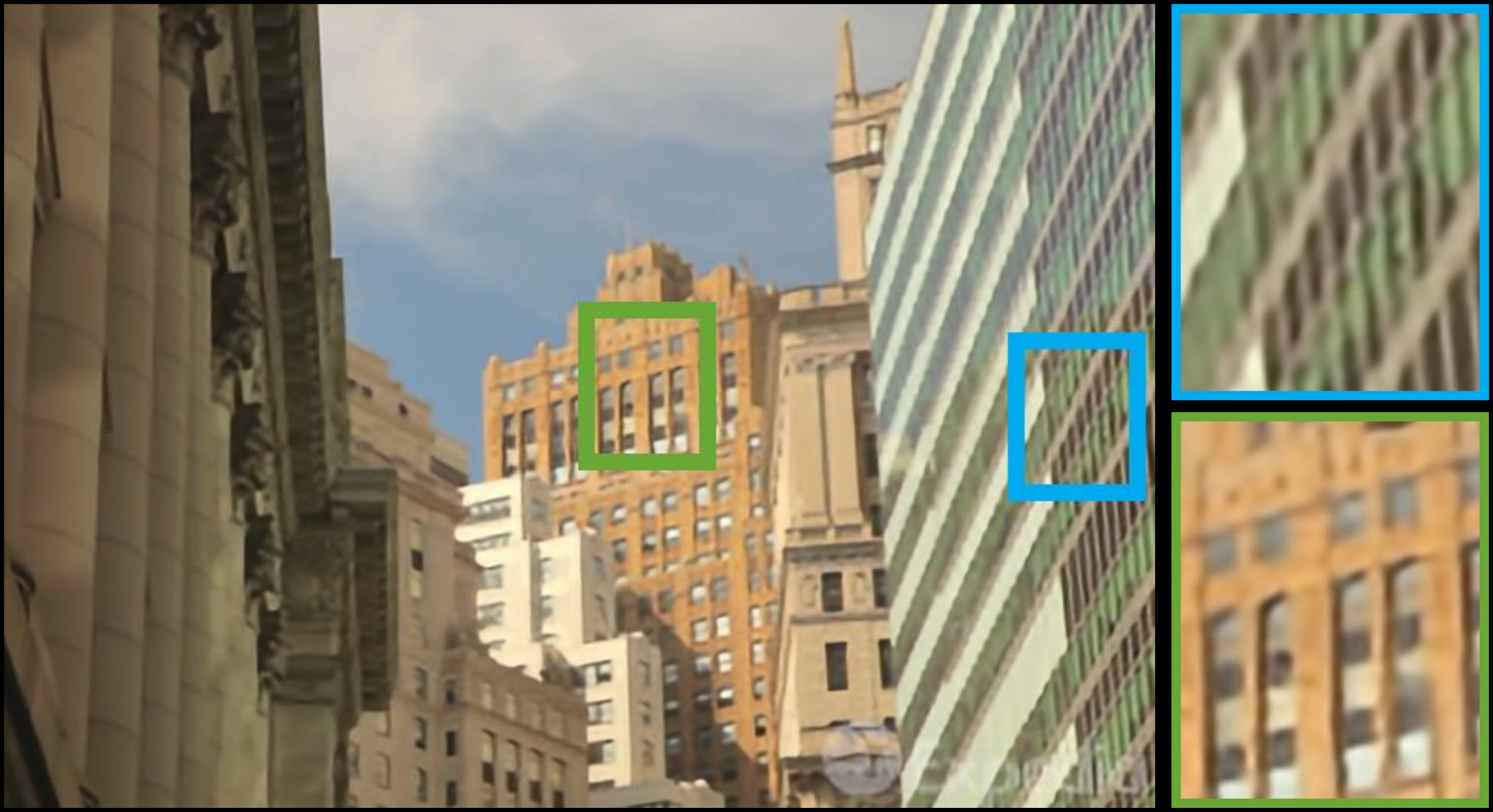
Ours

# Comparisons



Bicubic x4

# Comparisons

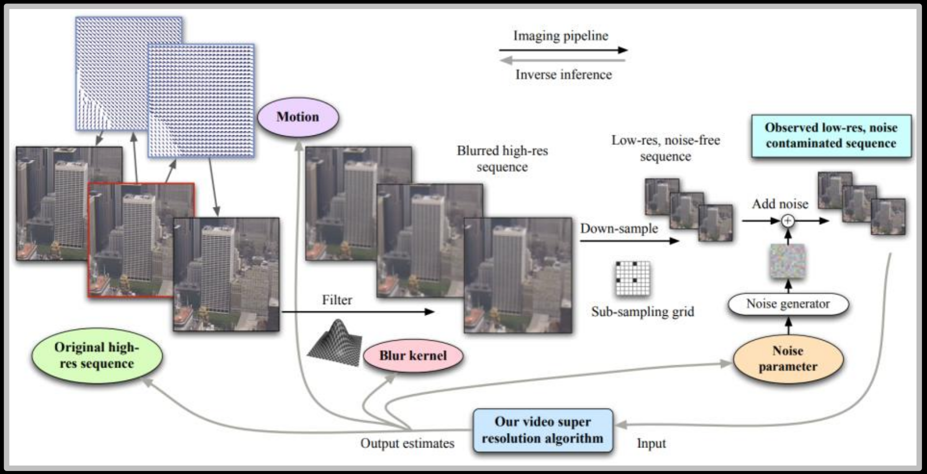


Ours

# Running Time

# Running Time

- BayesSR [Liu et al, 2011]



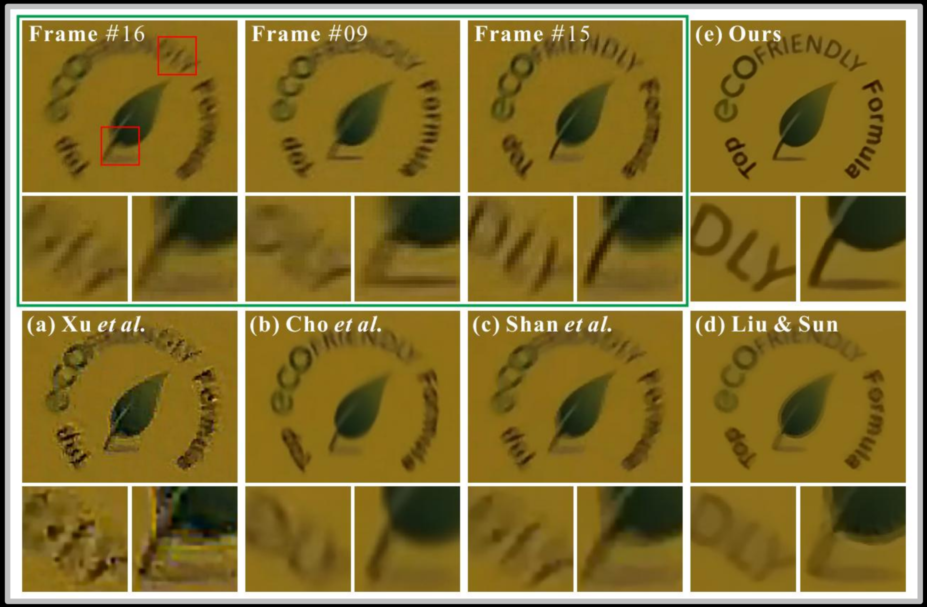

2 hour / frame

Frames: 31

Scale Factor: 4×

# Running Time

- MFSR [Ma et al, 2015]



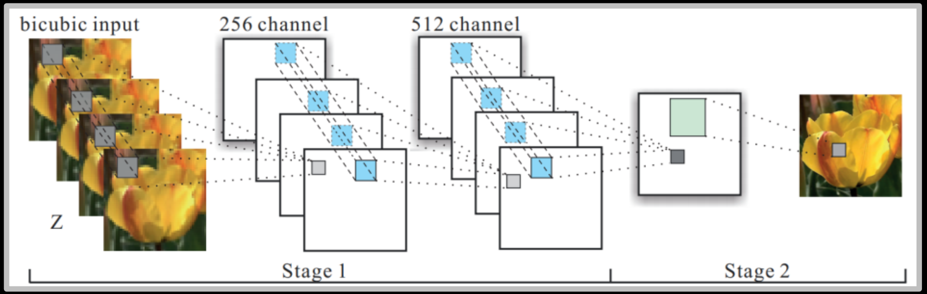

10 min / frame

Frames: 31

Scale Factor: 4×

# Running Time

- DESR [Liao et al, 2015]

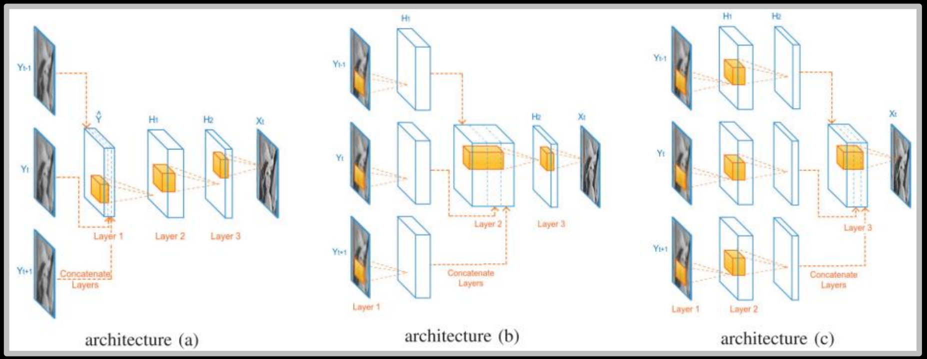



8 min / frame

Frames: 31
Scale Factor: 4×

# Running Time

- VSRNet [Kappeler et al, 2016]

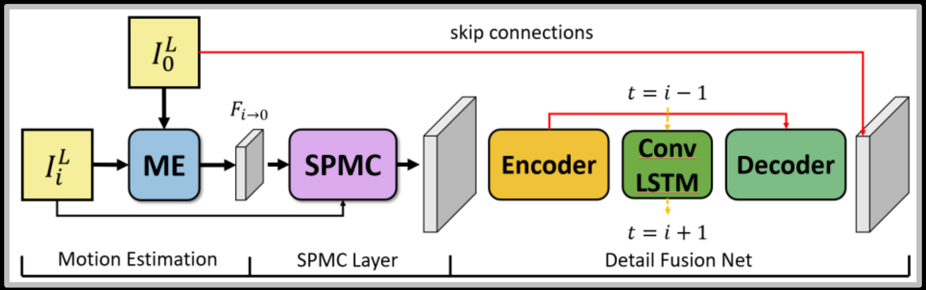



40 s / frame

Frames: 5
Scale Factor: 4×

# Running Time

- Ours (5 frames)



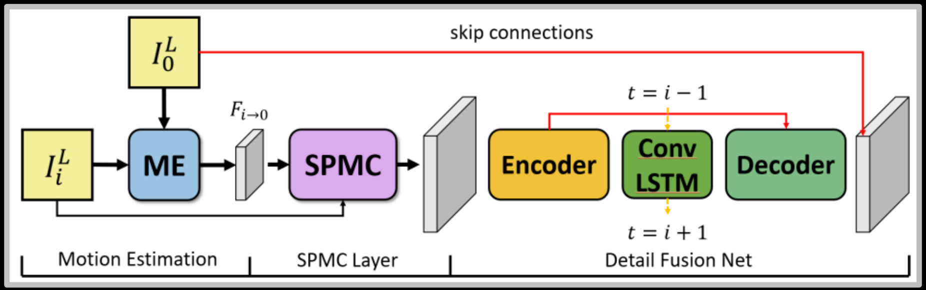

**0.19** s / frame

Frames: 5

Scale Factor: 4×

# Running Time

- Ours (3 frames)




**0.14** s / frame

Frames: 3
Scale Factor: 4×

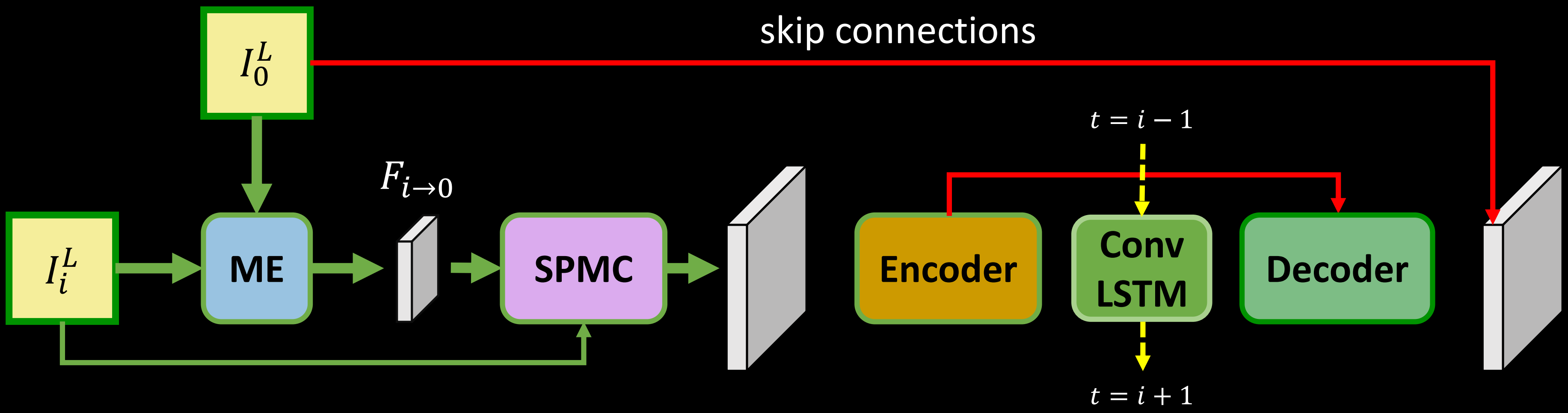
# More Results


Bicubic x4


Bicubic x4


Bicubic x4

# Summary

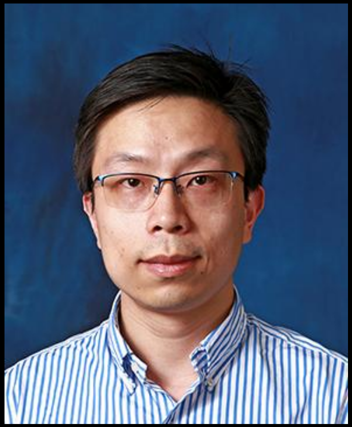



- End-to-end & fully scalable
- New SPMC layer
- High-quality & fast speed

# 腾讯优图X-Lab

- 成立于2017年，T5科学家领衔，负责前沿视觉技术研究与其它AI领域的融合

- 团队主要成员



贾佳亚 教授
杰出科学家

戴宇荣
专家研究员

沈小勇
高级研究员


INNOVATION
PROACTIVE
Integrity

# 我们需要

- 高级计算机视觉研究员
  - 负责计算机视觉算法的研究与产品落地，负责指导团队成员与实习生
  - 要求：博士或者优秀硕士
- 计算机视觉工程师
  - 计算机视觉算法的优化以及在云端，移动端的落地
  - 要求：两年以上相关开发经验
- 实习生
  - 要求：对计算机视觉具有浓厚兴趣，工程算法能力强者优先考虑

# Thank You