

# 图像大作业（去模糊）论文综述&技术方案

石耕源 2015013219

冯玉彤 2015013202

谢运帷 2015013185

---

## 论文综述

经过我们对于文献的研究和学习，我们将会主要介绍其中3篇对我们的技术方案有所指导的文章，并概述其他的文章。

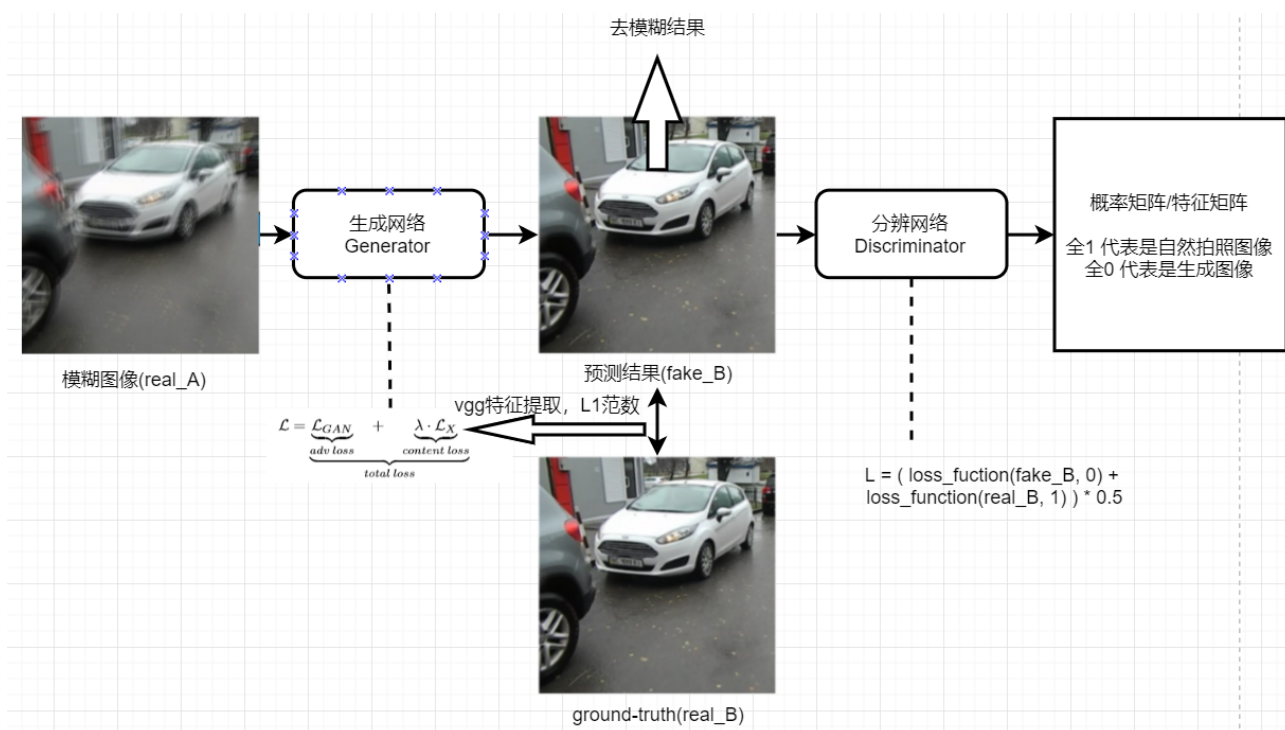
### Paper 1: DeblurGAN : Blind Motion Deblurring Using Conditional Adversarial Networks

这是推荐文献之中的第一篇，整体行文十分流畅，达到的效果也相当可圈可点。它使用了生成对抗网络来实现针对未知模糊核的图像的恢复。他们这样做主要是认为去模糊是一种图像到图像的转换，而这种转换，例如在风格迁移、分辨率扩大等问题上，GAN都有比较好的表现。除此之外，作者使用的Loss函数也是有多部分组成，既包含判断网络给出的结果，也包含去模糊后图像与ground-truth的差距。这样的设计在实践中效果更好。

这篇文章有三个突出的贡献：第一，它的处理速度很快，比当今的去模糊方法快了5倍以上；第二，它提出了一种生成数据的方法：模拟相机的运动，从清晰图片生成运动模糊图片，并将这种人工数据放入训练集中，可以加强去模糊的效果；第三，它提出了一种新的评价去模糊效果的评价函数，而这个评价函数是基于物体检测结果的。

---

我们来看一下这篇文章提出的生成对抗网络的结构：

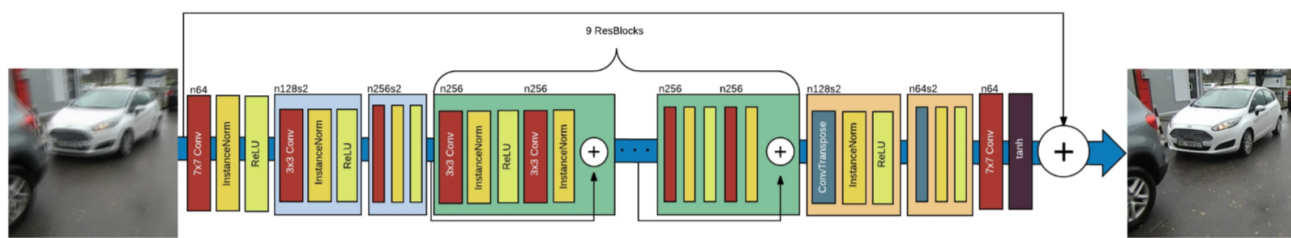


根据GAN的理论，整个网络在求解下式：

$$\min_G \max_D \mathbb{E}_{x \sim \mathbb{P}_r} [\log(D(x))] + \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [\log(1 - D(\tilde{x}))]$$

既分辨网络D可以分辨出自然拍摄的清晰图像和预测出的去模糊图像，而生成网络G可以生成和自然拍摄的清晰图像一样逼真的去模糊图像。这样，两者对抗的过程就可以使生成网络不断地进化，达到更好的去模糊效果。

整个网络的输入是模糊的拍摄图像，经过生成网络G得到了一个去模糊预测结果，随后又输入到分辨网络D中进行分辨。两个网络的优化过程是分开的，两者使用不同的Loss进行反向传播。生成网络的Loss定义，可以大致概括成这样，它包括一部分是：预测结果经过分辨网络D之后和1之间的距离，这反应了生成网络生成的图像有多逼真（越逼真代表越清晰）；另一部分是预测结果和ground-truth之间的内容距离，它定义为VGG19得到的feature\_map之间的L2距离。而分辨网络D的Loss则定义为预测结果经过分辨网络之后和0的距离，以及ground-truth经过分辨网络之后和1的距离的均值，这样当分辨网络D的Loss下降时，既说明了分辨网络D可以分出预测图像和ground-truth图像。



上图就是生成网络G的具体结构，它由多个ResBlock堆叠而成。可以较好的提取特征和进行预测。

这样，整个网络的结构和训练过程就定义好了。现在缺乏的就是输入real\_A的模糊图片和ground-truth real\_B的对应组合了。作者提到，可以使用GoPro拍摄的数据集，其中2103组模糊和清晰图片的对应组；或者是Kohler数据集，其中有4幅图片，分别被12个模糊核影响。但是，这些数据集并不满足所有的需求——在很多场景下我们并没有模糊图片和清晰图片的对应。作者基于前人工作，总结出了新的模糊图片生成方法。

模糊图片的生成，前人提出了一些方法，诸如：高速相机连续两张图像的平均、用线性运动对应的模糊核进行模糊操作、用若干采样点模拟曲线来创造模糊核等等。本文更进一步，同样是采取随机点拟合曲线的方法来创造模糊核，它采的每一个点，是由上一个点的位置与速度，以及高斯扰动、强制扰动和惯性所决定的。文中也给出了这一步的算法。

---

```

1: procedure BLUR(Img, M, Lmax, ps)
2:    $v_0 \leftarrow \cos(\phi) + \sin(\phi) * i$ 
3:    $v \leftarrow v_0 * L_{max} / (M - 1)$ 
4:    $x = \text{zeros}(M, 1)$ 
5:   for  $t = 1$  to  $M - 1$  do
6:     if  $\text{randn} < p_b * p_s$  then
7:        $\text{nextDir} \leftarrow 2 * v * e^{i * (\pi + (\text{randn} - 0.5))}$ 
8:     else:
9:        $\text{nextDir} \leftarrow 0$ 
10:     $dv \leftarrow \text{nextDir} + p_s * (p_g * (\text{randn} + i * \text{randn}) * I * x[t] * (L_{max} / (M - 1)))$ 
11:     $v \leftarrow v + dv$ 
12:     $v \leftarrow (v / \text{abs}(v)) * L_{max} / (M - 1)$ 
13:     $x[t + 1] \leftarrow x[t] + v$ 
14:   $\text{Kernel} \leftarrow \text{sub pixel interpolation}(x)$ 
15:   $\text{Blurred image} \leftarrow \text{conv}(\text{Kernel}, \text{Img})$ 
16:  return Blurred image

```

---

实验部分，本文使用了三个数据集进行实验：

第一个是GoPro数据集的实验，在测试集上，本文的方法体现出了较好的峰值信噪比，比其他方法更好的结构相似性，从人眼的视觉上来讲结果也更好。

	Sun <i>et al.</i>	Nah <i>et al.</i>	Xu <i>et al.</i>	DeblurGAN		
Metric	[36]	[25]	[44]	<i>WILD</i>	<i>Synth</i>	<i>Comb</i>
PSNR	24.6	28.3/29.1*	25.1	27.2	23.6	28.7
SSIM	0.842	0.916	0.89	0.954	0.884	0.958
Time	20 min	4.33 s	13.41 s	0.85 s		

在第二个Köhler的数据集上，结果表现也和GoPro数据集上的结果类似。

在第三个YOLO的数据集上，本文通过自己的模糊图片生成方法生成了多组街景和模糊街景的对照图片对，并通过自己的预测网络进行去模糊操作。再利用YOLO上一些预训练好的物体识别网络，来对比去模糊前和去模糊后不同的物体识别率，进而反应去模糊操作是否有效。

Method	prec.	recall	F1 score
no deblur	0.821	0.437	0.570
Nah <i>et al.</i> [25]	<b>0.834</b>	0.552	0.665
DeblurGAN WILD	0.764	0.631	0.691
DeblurGAN synth	0.801	0.517	0.628
DeblurGAN comb	0.671	<b>0.742</b>	<b>0.704</b>

上述给出了DeblurGAN之后，去模糊图片的一些物体识别评价价值。可以发现，DeblurGAN方法可以获得更好的recall和F1 score。本文评价这种利用物体识别方法来评价去模糊效果更好，更能反应人类的视觉效果，是否一个物体被还原的很清晰。

## Paper 3. Dynamic Scene Deblurring Using Spatially Variant Recurrent Neural Networks

### Background

本文主要针对的是相机震动和物体运动产生的模糊，是基于CNN的盲去模糊方法。文中使用RNN实现去卷积过程，并且针对现有CNN方法需要计算庞大的模型的问题提出解决方案，最终评价指标表现优秀。

### Contribution

本文的贡献主要有三：

- 去卷积 RNN 网络的权重由 CNN 网络学习得到，最终产生端对端可训练的模型，能够应对多样化的模糊情况
- 将去模糊过程阐述为求解无限冲击响应 (Infinite Impulse Response, IIR) 模型问题，指出用 RNN 可以有效求解该问题，并具有较大感受野
- 实验表明该方法的精确度、速度、模型大小都优于最先进方法

### Method

图像模糊的过程可以表示为

$$y[n] = \sum_{m=0}^M k[m]x[n-m] \quad (1)$$

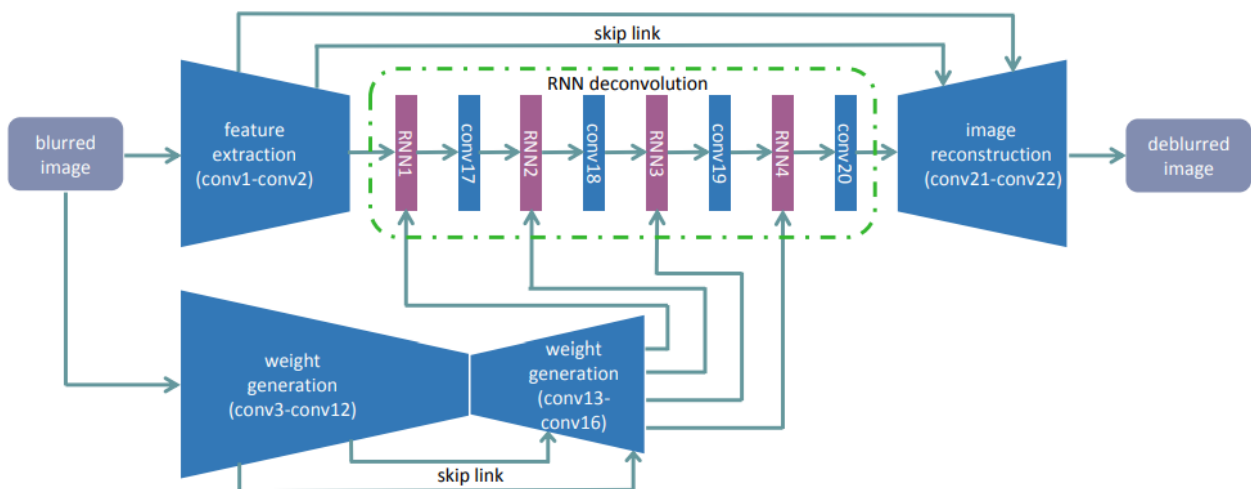
其中  $x$  表示原始图像， $k$  表示模糊核， $y$  表示模糊后图像， $M$  表示模糊核大小。可反解原图像  $x$  为

$$x[n] = \frac{y[n]}{k[0]} - \frac{\sum_{m=1}^M k[m]x[n-m]}{k[0]} \quad (2)$$

$x[n]$  的表达中又含有序列前  $M$  时刻的值。该方程可表示为信号处理领域的 IIR 模型。方程展开后可表示为对模糊信号  $y$  作用逆滤波，而实验表明，逆滤波矩阵的参数比原模糊核的参数多很多，因此在使用基于 CNN 的去卷积方法时，需要使用较大的模型来提高感受野的范围。

同时注意到 (2) 中实际上只有  $M$  个参数需要求解，即  $k[m]$ 。使用 RNN 可以以较小的模型求解该问题。同时还需要在每个 RNN 网络之后加入 CNN 以融合不同方向学到的信息。

本文设计的网络结构如下。



先于 feature extraction 阶段使用卷积层提取原图像的 feature map，后通过 weight generation 部分生成 RNN 网络的权重。将 feature map 通过四个 RNN 模块完成去卷积后重建图片并进行评价。

该网络结构的主要特点有：

- 每个 RNN 模块都考虑了四个方向，并在其后加入一个卷积层来聚合输出，提供了较大的感受野
- weight generation 部分除了最后一个之外的每个卷积层后面使用 ReLU 整流，最后一个卷积层后面使用双曲正切函数 (tanh)，从而使权值分布在 [0, 1] 区间
- feature extraction, RNN deconvolution, reconstruction 部分除了最后一个之外的每个卷积层后面加使用 负斜率为0.1的 Leaky ReLU 整流
- 重建部分使用双线性插值上采样，防止 grid artifact 问题
- 使用 skip link 以加快训练速度和防止梯度消失

## Experiment

本文在GOPRO数据集<sup>1</sup>上进行了实验，与传统方法<sup>2 3 4</sup>，基于 CNN 的方法<sup>5 6 1</sup>进行了对比。

各方法的 PSNR (尖峰信噪比) 值、SSIM (结构相似性) 值如下表所示。可见本文方法均优于其他方法。

	Whyte	Xu	Sun	Pan	Liu	Nah	Gong	Proposed
PSNR	24.5	20.3	25.3	23.5	25.7	28.5	26.1	<b>29.2</b>
SSIM	0.85	0.74	0.85	0.83	0.86	0.91	0.86	<b>0.93</b>
time(s)	700	3800	1500	2500		15	1500	<b>1.4</b>
size(MB)			54.1			303.6	41.2	<b>37.1</b>

在真实模糊数据集<sup>7</sup>上，应用本文提出的方法生成的图像也比其他方法更加清晰。

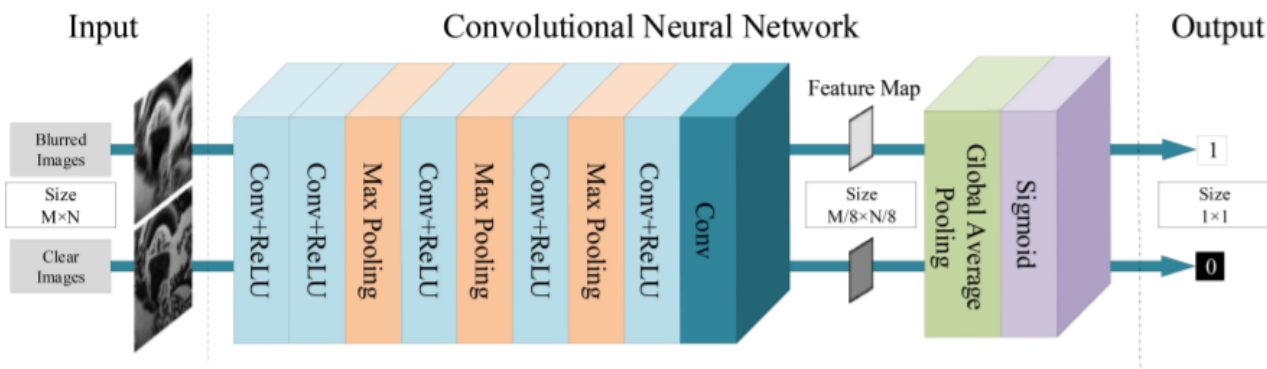
## Paper 4 : Learning a Discriminative Prior for Blind Image Deblurring

阅读此篇论文遇到了很大的困难，因为此论文的去模糊部分运用到了很多的数学，所以就简述一些整体的思路。

$$\min_{I,k} \|I \otimes k - B\|_2^2 + \gamma \|k\|_2^2 + \mu \|\nabla I\|_0 + \lambda f(I), \quad (4)$$

在一个盲的去模糊问题中，一个常用的方法即是求解上述优化问题。作者将问题分为对于I和k的两个子问题，使用迭代下降的方法，求解两个子问题，就得到了原清晰图像I和模糊核k。在求解上述优化问题的过程中，有一项 $f(I)$ ，它表达的是这个此图像是否清晰。作者训练了一个CNN网络来计算这个 $f(I)$ 。

这个分辨网络如下：



这个网络接受一张图片，输出1（模糊）或者0（清晰）。

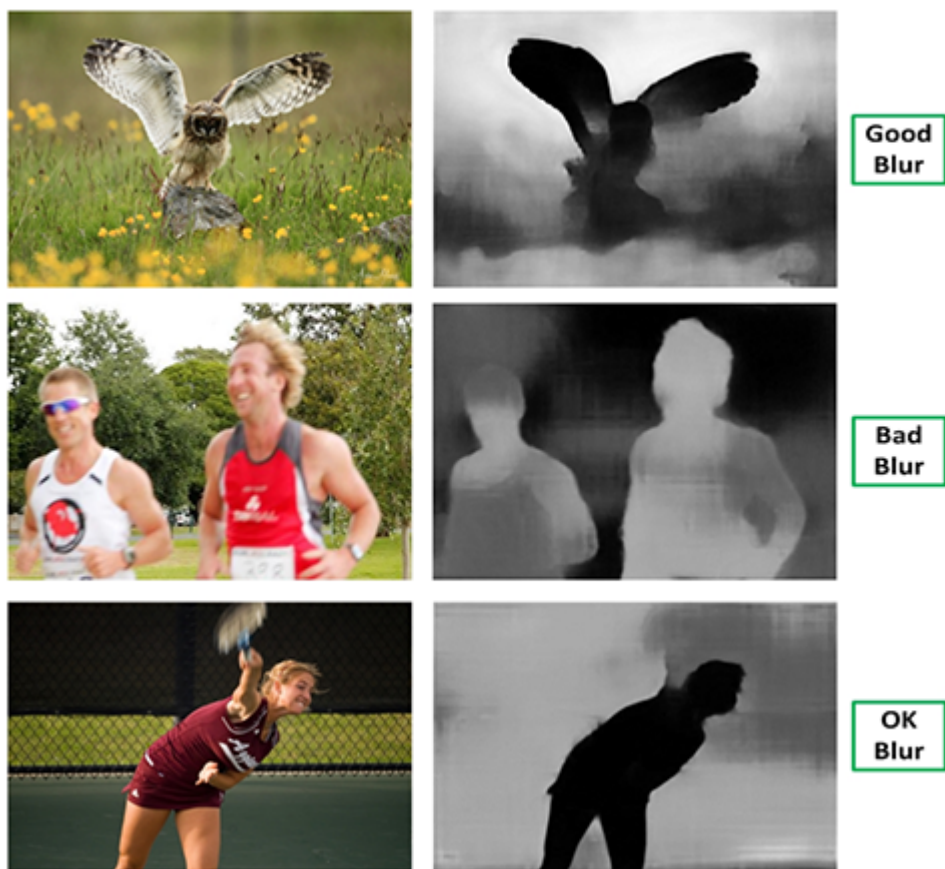
作者主要在自己的随机生成的模糊核影响后的模糊图片上进行了实验，相比方法，它可以更准确的预测模糊核，达到了更好的效果。在现实拍摄的模糊图片上，也相对表现不错。

## Paper 5. Learning to Understand Image Blur

### Background

在图像去模糊的相关工作中，大部分研究主要对图像中的模糊进行估计与去除，但忽视了模糊需求度 (blur desirability) 的影响，即图像中的模糊既可能是由于手抖、对焦失败等原因产生的负面模糊，也可能是被主动对焦虚化的正面模糊，如下图所示，可大致将图像的模糊需求度分为好、坏和一般。去模糊工作应该先判断图像中存在的模糊是哪一种类型，再由此进行进一步的去除操作，本篇论文研究和提供了对图片中模糊的需求度的理解与估计方法。





## Contributions

本篇论文的贡献主要是以下 3 点：

- 提出了端到端的网络 ABC-FuseNet，对注意力图 (Attention map)、模糊图 (Blur map) 和内容特征图 (Content feature map) 三类特征进行融合，并给出整张图的模糊需求度估计结果；
- 建立了大规模数据集 SmartBlur，包含了不同原因导致的模糊图片数据，并提供了每张图像的逐像素模糊度和整幅图模糊需求度；
- 在 SmartBlur 和另一个模糊图像数据集上进行了实验验证，表明了其提出的方法在模糊估计和模糊需求度估计上的性能超越了当前最好的方法。

## The SmartBlur Dataset

由于现有数据集中缺少对模糊需求度的标注，本篇论文的作者首先建立了数据集 SmartBlur，其中图像的标注主要为以下两种：

- 模糊度：该标注为逐像素级别，表示该位置处的模糊程度，本数据集将其分为 High Blur, Medium Blur, Low Blur 和 Clear 四种类别；
- 模糊需求度：该标注为逐图像级别，对存在模糊的图像，给出该图像中的模糊是正面还是负面的，本数据集将其分为 Bad-Blur, Ok-Blur, Good-Blur 和 No-Blur 四种类别。

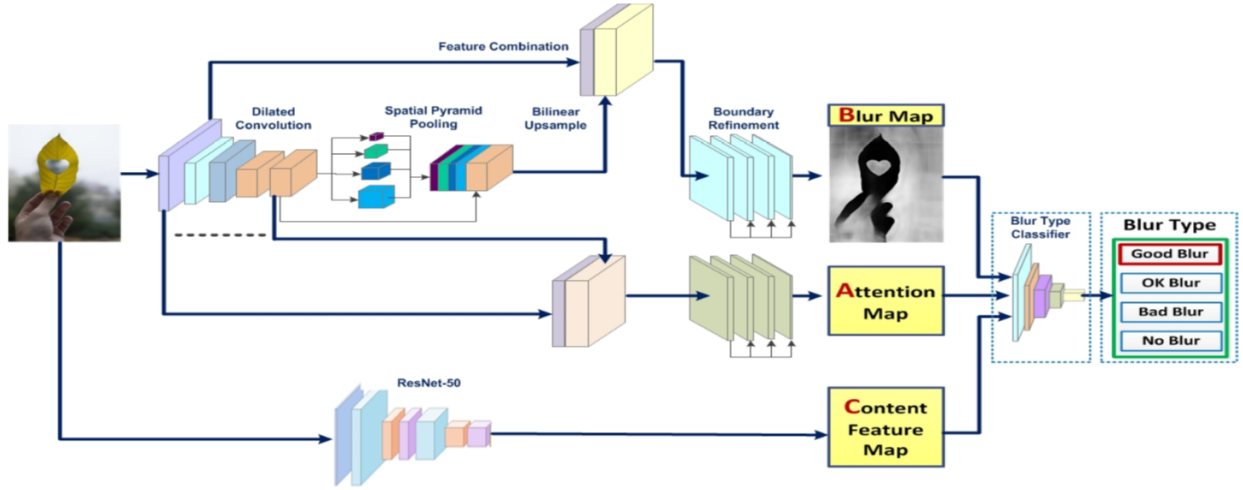
此外，本数据集在建立过程中特别注意了不同模糊原因及不同模糊需求度的图像所占比例，保证了数据集的平衡性。

## Method

本篇论文提出的 ABC-FuseNet 的网络结构如下图所示，主要分为四部分：

- 利用扩张全卷积神经网络、空间金字塔池化、边界校正，生成图像的模糊图；

- 利用全卷积神经网络，生成图像的注意力图；
- 利用用于图像识别的卷积神经网络，生成图像的内容特征图；
- 综合以上得到的 3 种特征图，利用模糊类型分类器，得到最终的模糊类型/模糊需求度。



图像的模糊图代表每个像素位置的模糊度，生成模糊图任务的主要困难是，现有的卷积神经网络存在大量池化下采样操作，削弱了对像素位置和小范围的感知能力，此外模糊区域的大小不固定，而卷积神经网络同一层的感受野大小是相同的。本方法在图像识别网络 Inception-V2 的基础上，一方面去除了原有的下采样操作，同时利用扩张卷积替换了原有的卷积操作；另一方面通过空间金字塔池化，将不同维度的特征图池化为统一尺寸，又拼接了来自高维度特征图的上采样结果后，进行一次双线性插值的上采样操作。通过这些改进，保留了特征提取过程中来自低维度特征图的位置信息，融合了不同感受野区域下的特征提取结果。为了得到更好的边界，会再次拼接低维度特征并输入边界校准层，得到最终的模糊图。

图像的注意力图代表在语义层面对图像不同位置的注意程度，内容特征图则代表对图像中具体内容的理解，为了帮助判断图像中模糊的好坏，需要先提取出以上两种特征图进行辅助，本方法中均适用预训练模型实现了两种特征图的提取。

在得到三种特征图后，需要使用一个融合网络接收三种特征并给出模糊需求度的估计结果，本方法使用了一个简单且轻量级的卷积神经网络实现这一功能，为了更好地进行特征融合，本方法没有直接将其进行逐频道拼接，而是使用了一种双重注意力机制，设三类特征图分别为  $A_m, B_m, C_m$ ，融合特征为  $B_m A_m, B_m(1 - A_m), C_m$ 。

本方法的网络模型中需要进行训练的部分为模糊图生成和模糊需求度估计，其中模糊需求度估计采用 Softmax Cross-Entropy Loss，标记为  $L_{Bm}$ ；模糊图生成的损失函数设计如下

$$L_{Bm} = \frac{1}{2N} \sum_{i=1}^N \sum_{p=1}^P \| \frac{1}{1 + \exp(-b_i(p; \Theta))} - b_i^0(p) \|_2^2, \quad (3)$$

其中， $b_i(p; \Theta)$  为第  $i$  张图中第  $p$  个像素点的估计模糊度， $\Theta$  为估计网络的参数， $b_i^0(p)$  为第  $i$  张图中第  $p$  个像素点的实际模糊度。

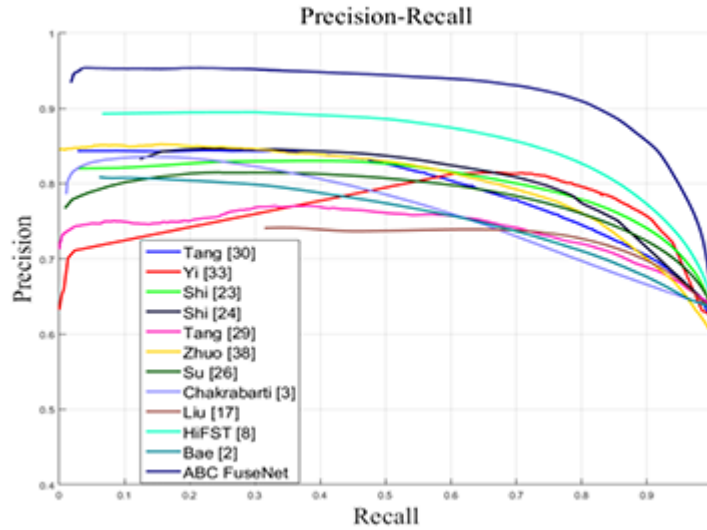
最终的损失函数为：

$$L = L_{Bm} + \lambda L_{Bc} \quad (4)$$

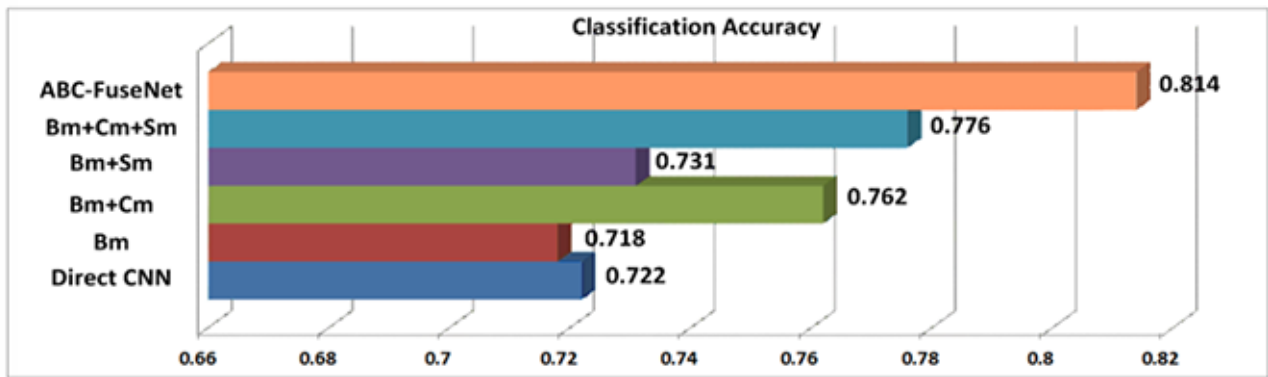
## Experiments

本篇论文首先在公开数据集 CUHK 上对其模糊图生成部分进行了实验验证，该数据集包含 1000 张标注了模糊区域的图像，标注类别为模糊/不模糊，实验对比了一系列模糊图生成算法的 Precision-Recall 结果，如下图所示，超越了现有的最优算法。





接下来在自己提出的 SmartBlur 数据集上进行了模糊图生成和模糊需求度估计的实验，其中模糊图生成结果同样与上述算法进行了对比并达到了最优结果；在模糊需求度估计方面，其分类精度达到了 81.4%，同时还进行对比实验验证了其双重注意力机制的有效性，如下图所示，对比了不同融合方法下的分类精度，可以看到最终采用的融合方式达到了最优结果。



## Paper 6 : Non-blind Deblurring: Handling Kernel Uncertainty with CNNs

此篇意在解决一个重要的问题，在现实中，模糊核往往是有噪声的。在现有的非盲的去模糊方法中，都需要估计出一个准确的模糊核，才能保证最终恢复出的图像清晰。非盲的方法，使用一个MAP最大后验概率估计的方法，通过估计出的模糊核，来估计清晰的图像：

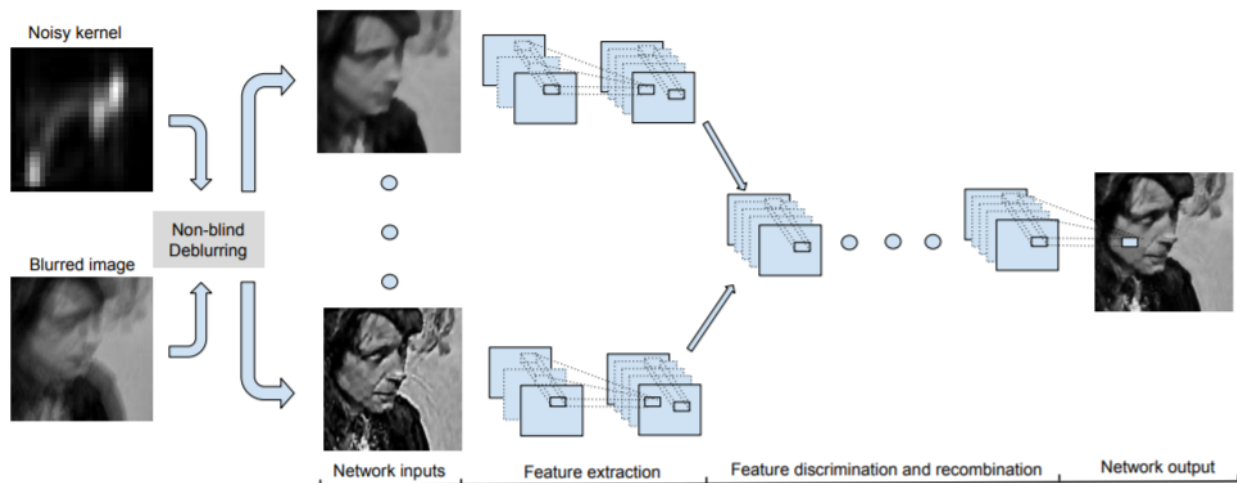
$$\hat{l} = \arg \min_l \frac{\lambda}{2} \|l * \hat{k} - b\|^2 + P(l)$$

这样的方法，不光依赖我们选取的图像范数 $P(l)$ ，还依赖这个超参数 $\lambda$ 。作者给出了一组实验，证明了不同的 $\lambda$ 和 $P(l)$ 下，我们恢复的图像的效果不同。



Figure 1. Impact on the restored image quality with varying  $\lambda$ . (a) Latent image. (b) Ground truth kernel. (c) Blurred image generated using (a) and (b). (d) Noisy kernel obtained from [32]. Estimated images using kernel in (d) with  $\lambda_{2/3} =$  (e)  $2e4$ , (f)  $2e3$ , and (g)  $2e2$ , with  $\lambda_2 =$  (h)  $1e4$ , (i)  $1e3$ , and (j)  $1e2$ , and with  $\lambda_p =$  (k)  $64e5$ , (l)  $64e4$ , and (m)  $64e3$ . (n) Average MSE of restored images for varying  $\lambda$  values.

在不同的 $\lambda$ 下，利用有噪声的模糊核还原出的图片，有些细节更丰富却有人工的痕迹，有些细节不丰富但人工痕迹少。在这样的结果下，作者提出一种利用CNN把不同的还原结果综合一下，等到一个良好的去模糊结果的方法：



先通过不同超参数进行预测，得到不同的还原结果，在通过CNN网络进行提取特征，把特征组合起来，最终在换成一张图像，这张图像就是最终的还原结果，既保留了细节，有不会使其看起来有很多人工痕迹（纹理）。

## Extra: Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring

### Background

本文<sup>1</sup>是CVPR 2017的论文，提供了自然模糊数据集，用于测试基于CNN的盲去模糊方法效果。现有的非盲去模糊算法往往基于对模糊核的预测，而预测结果的不精确会严重影响生成的图片。而由于有监督学习的真实模糊数据集的缺乏，现有CNN算法只能处理有限的特定模糊问题。

### Contribution

本文的主要贡献有三：

- 提出的CNN网络不预测或假定模糊核的信息，因此不会出现因预测模糊核偏差而导致的失真
- 计算多尺度 Loss，使网络更快收敛；计算对抗 Loss 改进效果

- 提供了具有清晰图片的数据集，且不依赖模糊核

## Dataset

文中提出的数据集 GOPRO 中的模糊是由物体移动、相机震动等多种因素的平均生成的。模糊过程由下述公式给出：

$$B = g \left( \frac{1}{M} \sum_{i=0}^{M-1} S[i] \right) \quad (5)$$

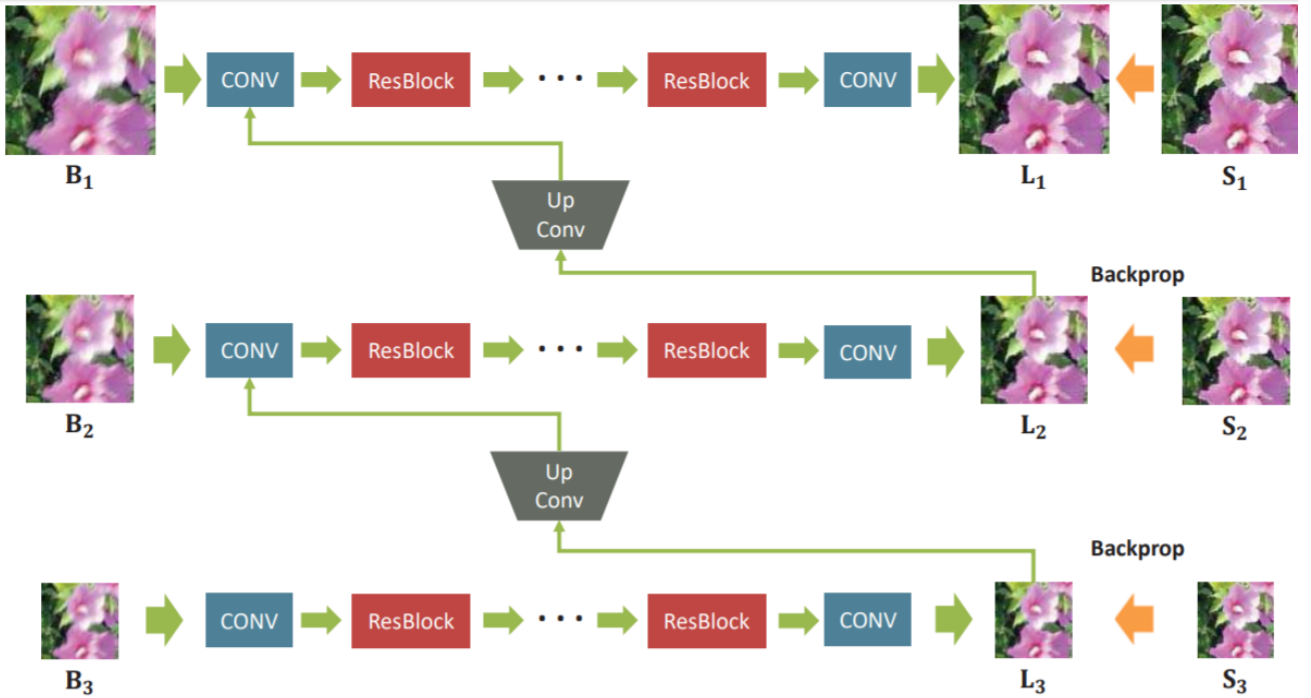
其中  $M$  表示采样范围。 $g$  函数表示 CRF(Camera Response Function)，即由原图像得到观测图像的映射。实践中无法获知该 CRF 的准确值，使用常用的  $\gamma = 2.2$  曲线，它代表了已知 CRF 的平均值，因此有映射

$$g(x) = x^{\frac{1}{\gamma}} \quad (6)$$

用不同的  $M$  来模拟相机聚焦的时间。

## Method

本文的网络结构采用了高斯金字塔模型，完整网络如下图所示。



借鉴了残差网络的结构，对原始输入图片每次进行下采样，再将高层的反馈回传到低层网络中，最终采用原始尺寸图片的输出作为全部网络的输出，文章认为这样可以同时保留图片的高层与底层特征。

训练采用了标准的 MSE Loss 和最近提出的对 Adversarial Loss，其算法为：

$$L_{cont} = \frac{1}{2K} \sum_{k=1}^K \frac{1}{c_k w_k h_k} \|L_k - S_k\|^2 \quad (7)$$

$$L_{adv} = \mathbb{E}_{S \sim P_{sharp}(S)} [\log D(S)] + \mathbb{E}_{B \sim P_{blurry}(B)} [\log(1 - D(G(B)))] \quad (8)$$

最终的 Loss 取二者的加权平均。

实验在本文提出的 GOPRO 数据集上进行，结果表明效果显著优于现有方法。

# 技术方案

经过讨论，我们将采用DeBlurGAN的整体结构，并且在它的基础上进行一些改进：

- 1、我们发现DeBlurGAN生成的图像的PSNR值相对低一些，而在调研中，RNN网络的文章生成的图像结果PSNR较好。我们考虑将生成网络G的部分进行修改和替换，进行对比实验。
- 2、我们发现去模糊得到的结果局部还是有些模糊，也有一些纹理问题，我们考虑使用滤波器来进一步消除这些细节问题
- 3、DeBlurGAN的判别网络D有些复杂，借鉴其他的文章2值的判别输出，我们会尝试对于此部分进行改进
- 4、尝试在野生的无对照数据集上进行训练和实验

我们整体的工作流程大概如下：

- 1、搜集数据集
- 2、编写新的生成网络model
- 3、编写新的判断网络
- 4、训练并实验
- 5、总结&展示

---

## 参考文献

- 
1. Nah, Seungjun, Tae Hyun Kim, and Kyoung Mu Lee. "Deep multi-scale convolutional neural network for dynamic scene deblurring." *CVPR*. Vol. 1. No. 2. 2017. [↩↩↩](#)
  2. L. Xu, S. Zheng, and J. Jia. Unnatural l0 sparse representation for natural image deblurring. In *CVPR*, 2013. [↩](#)
  3. J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Blind image deblurring using dark channel prior. In *CVPR*, 2016. [↩](#)
  4. O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. *IJCV*, 2012. [↩](#)
  5. J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *CVPR*, 2015. [↩](#)
  6. D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. v. d. Hengel, and Q. Shi. From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. In *CVPR*, 2017. [↩](#)
  7. S. Cho, J. Wang, and S. Lee. Video deblurring for hand-held cameras using patch-based synthesis. *TOG*, 2012. [↩](#)