

5 月 21 日-5 月 29 日工作汇报

Ku Jui

May 2023

Contents

1	文献阅读	1
1.1	Pre-knowledge	1
1.1.1	loss function for CV	2
1.2	Paper reading	13
1.2.1	Switching gaussian mixture variational rnn for anomaly detection of diverse cdn websites [1]	13
2	个人工作进展	15
2.1	关于“华为杯”的一点思考	15
2.1.1	适用多场景的通用化时序预测算法	15
2.2	梳理损失函数	16
2.3	复现 KinD 代码	17
2.3.1	Requirement	17
2.3.2	Model	17
3	下周工作计划	21

1 文献阅读

1.1 Pre-knowledge

模型中的损失函数是用来衡量模型的预测值和真实值 (Ground Truth) 之间的差异程度的函数，它可以反映模型的优化方向和性能指标¹。它是一种衡量模型预测结果与真实结

¹<https://zhuanlan.zhihu.com/p/375968083>

果之间差异的方法，用于指导模型参数的更新。在训练过程中，通过不断最小化损失函数来优化模型参数，使模型能够更好地拟合数据²。因此，需要使用合适的损失函数，当模型在数据集上进行训练时，该函数可以适当地惩罚模型³。

不同的损失函数适用于不同的任务和数据分布，例如回归问题常用的有均方误差损失函数 (MSE，也叫做 \mathcal{L}_2 损失函数) 和 \mathcal{L}_1 损失函数，分类问题常用的有交叉熵损失函数 (Cross Entropy Loss) 等⁴。损失函数的选择会影响模型的收敛速度和精度，因此需要根据具体情况选择合适的损失函数⁵。

目前常用的损失函数是从相关视觉任务中借用的，但这些损失函数可能并不完全适用于低照度图像增强 LLIE。因此，需要设计更适合 LLIE 的损失函数，以更好地驱动深度网络的优化。这可以通过研究人类对图像质量的视觉感知来实现，使用深度神经网络来近似人类视觉感知，并将这些理论应用于损失函数的设计。损失函数可以分为两个大类：回归问题和分类问题。

1.1.1 loss function for CV

\mathcal{L}_1 -loss

平均绝对误差 (MAE) 损失，也称 \mathcal{L}_1 范数损失，计算实际值和预测值之间绝对差之和的平均值。

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^N \|\hat{y}_i - y_i\|_1 \quad (1)$$

适用于回归问题，MAE loss 对异常值更具鲁棒性，尤其是当目标变量的分布有离群值时 (小值或大值与平均值相差很大)。

函数: `torch.nn.L1Loss`

\mathcal{L}_2 -loss

均方误差 (MSE) 损失，也称为 \mathcal{L}_2 范数损失，计算实际值和预测值之间平方差的平均值⁶。

²<https://zhuanlan.zhihu.com/p/436809988>

³<https://zhuanlan.zhihu.com/p/473113939>

⁴https://blog.csdn.net/weixin_57643648/article/details/122704657

⁵https://www.zhihu.com/tardis/zm/art/136047113?source_id=1005

⁶<https://blog.csdn.net/yanyuxiangtoday/article/details/119788949>

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^N \|\hat{y}_i - y_i\|_2^2 \quad (2)$$

平方意味着较大的误差比较小的误差会产生更大的惩罚，所以 $\mathcal{L}_2 - loss$ 的收敛速度要比 $\mathcal{L}_1 - loss$ 要快得多。但是， $\mathcal{L}_2 - loss$ 对异常点更敏感，鲁棒性差于 $\mathcal{L}_1 - loss$ 。

$\mathcal{L}_1 - loss$ 损失函数相比于 $\mathcal{L}_2 - loss$ 损失函数的鲁棒性更好。因为以 $\mathcal{L}_2 - loss$ 范数将误差平方化 (如果误差大于 1，则误差会放大很多)，模型的误差会比以 $\mathcal{L}_1 - loss$ 范数大的多，因此模型会对这种类型的样本更加敏感，这就需要调整模型来最小化误差。但是很大可能这种类型的样本是一个异常值，模型就需要调整以适应这种异常值，那么就导致训练模型的方向偏离目标了⁷。

对于大多数回归问题，一般是使用 $\mathcal{L}_2 - loss$ 而不是 $\mathcal{L}_1 - loss$ ， $\mathcal{L}_2 - loss$ 在 LLIE 中，具有的相当很有效的恢复效果，但是由于平方项，它通常会产生模糊的恢复结果，如 [2] 所述，平方项为相对较小的误差提供了较低的梯度。。

函数: `torch.nn.MSELoss`

Regularization of the \mathcal{L}_1 and \mathcal{L}_2 loss functions

正则化的基本思想是通过在损失函数中加入额外信息，以便防止过拟合和提高模型泛化性能。无论哪一种正则化方式，基本思想都是希望通过限制权重的大小，使得模型不能任意拟合训练数据中的随机噪声，正则化实际是在损失函数中加入刻画模型复杂程度的指标⁸。

对应的 L1 正则损失函数:

$$\mathcal{L}_{norm1} = \mathcal{L}_1(\hat{y}, y) + \lambda \sum_{\omega} \|\omega\|_1 \quad (3)$$

对应的 L2 正则损失函数:

$$\mathcal{L}_{norm2} = \mathcal{L}_2(\hat{y}, y) + \lambda \sum_{\omega} \|\omega\|_2^2 \quad (4)$$

假设 $\mathcal{L}_1(\hat{y}, y)$ 和 $\mathcal{L}_2(\hat{y}, y)$ 是未加正则项的损失， λ 是一个超参，用于控制正则化项的大小，惩罚项 ω 用于惩罚大的权重，隐式地减少自由参数的数量。

正则化是如何降低过拟合现象的？

正则化之所以能够降低过拟合的原因在于，正则化是结构风险最小化的一种策略实现。给损失函数加上正则化项，能使得新得到的优化目标函数 $h = f + normal$ ，需要在 f 和

⁷<https://zhuanlan.zhihu.com/p/137073968>

⁸https://blog.csdn.net/weixin_41960890/article/details/104891561

$normal$ 中做一个权衡 (trade-off), 如果还像原来只优化 f 的情况下, 那可能得到一组解比较复杂, 使得正则项 $normal$ 比较大, 那么 h 就不是最优的, 因此可以看出加正则项能让解更加简单, 符合奥卡姆剃刀理论, 同时也比较符合在偏差和方差 (方差表示模型的复杂度) 分析中, 通过降低模型复杂度, 得到更小的泛化误差, 降低过拟合程度⁹。

PyTorch 实现: \mathcal{L}_2 正则项是通过 optimizer 优化器的参数 `weight_decay(float, optional)` 添加的, 用于设置权值衰减率, 即正则化中的超参 λ , 默认值为 0。

```
1 optimizer = torch.optim.SGD(model.parameters(), lr=0.01, weight_decay=0.01)
```

Smooth \mathcal{L}_1 loss function

Smooth \mathcal{L}_1 损失函数是由 Girshick R 在 Fast R-CNN 中提出的, 主要用在目标检测中防止梯度爆炸。它是一个分段函数, 在 $[-1, 1]$ 之间是 \mathcal{L}_2 损失, 其他区间就是 \mathcal{L}_1 损失。这样即解决了 \mathcal{L}_1 损失在 0 处不可导的问题, 也解决了 \mathcal{L}_2 损失在异常点处梯度爆炸的问题¹⁰。

$$\text{smooth } \mathcal{L}_1 \text{ loss} = \frac{1}{N} \sum_{i=1}^N \begin{cases} \frac{\|\hat{y}_i - y_i\|_2^2}{2\beta}, & \|\hat{y}_i - y_i\| < \beta, \\ \|\hat{y}_i - y_i\|_1 - \frac{1}{2}\beta, & \|\hat{y}_i - y_i\| \geq \beta. \end{cases} \quad (5)$$

一般取 $\beta = 1$ 。smooth \mathcal{L}_1 和 \mathcal{L}_1 -loss 函数的区别在于, smooth \mathcal{L}_1 在 0 点附近使用 \mathcal{L}_2 使得它更加平滑, 它同时拥有 \mathcal{L}_2 -loss 和 \mathcal{L}_1 -loss 的部分优点。

函数: `torch.nn.SmoothL1Loss`

Huber loss function

\mathcal{L}_2 -loss 但容易受离群点的影响, \mathcal{L}_1 -loss 对离群点更加健壮但是收敛慢, Huber Loss 则是一种将 MSE 与 MAE 结合起来, 取两者优点的损失函数, 也被称作 Smooth Mean Absolute Error Loss。其原理很简单, 就是在误差接近 0 时使用 \mathcal{L}_2 -loss, 误差较大时使用 \mathcal{L}_1 -loss

$$J_{Huber}(\delta) = \frac{1}{N} \sum_{i=1}^N \begin{cases} \frac{1}{2} \|\hat{y}_i - y_i\|_2^2, & \|\hat{y}_i - y_i\| < \delta, \\ \delta \left(\|\hat{y}_i - y_i\|_1 - \frac{1}{2}\delta \right), & \|\hat{y}_i - y_i\| \geq \delta. \end{cases} \quad (6)$$

⁹<https://zhuanlan.zhihu.com/p/35356992>

¹⁰<https://zhuanlan.zhihu.com/p/261059231>

残差比较小时, Huber Loss 是二次函数; 残差比较大时, Huber Loss 是线性函数 (残差, 即观测值和预测值之间的差值)。与 \mathcal{L}_2 -loss 相比, Huber 损失对数据中的异常值不那么敏感。使函数二次化的小误差值是多少取决于“超参数” δ , 它可以调整。当 $\delta = 1$ 时, 退化成 smooth \mathcal{L}_1 Loss。

函数: `torch.nn.HuberLoss`

log-MSE

$$J_{log-MSE} = 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^N \|\hat{y}_i - y_i\|_2^2 \right) \quad (7)$$

Perceptual loss function

感知损失 (Perceptual Loss) 是一种用于比较两个看起来相似的图像的损失函数, 这一损失函数由 Johnson et al. [3] 提出。它用于比较图像之间的高层次差异, 如内容和风格差异¹¹。它已被广泛用作图像合成任务 (包括图像超分辨率和风格转换) 中的有效损失项。

感知损失函数用于比较两个看起来相似的不同图像, 例如同一张照片, 但偏移了一个像素。该函数用于比较图像之间的高级差异, 例如内容和样式差异。感知损失函数与每像素损失函数非常相似, 因为两者都用于训练前馈神经网络以进行图像转换任务。感知损失函数是一个更常用的组件, 因为它通常提供有关风格迁移的更准确的结果。

简而言之, 感知损失函数的工作原理是将所有像素之间的所有平方误差相加并取平均值。这与每像素损失函数形成对比, 后者对像素之间的所有绝对误差求和 [3]。

作者认为感知损失函数不仅在生成高质量图像方面更准确, 而且在优化后也快了三倍。神经网络模型在图像上进行训练, 其中感知损失函数基于从已训练网络中提取的高级特征进行优化。

$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \|\phi_j(\hat{y}) - \phi_j(y)\|_2^2 \quad (8)$$

其中 \hat{y} 为输出图像, y 为目标图像, ϕ 为损失网络。 $\phi_j(x)$ 为处理图像 x 时损失网络 ϕ 的第 j 层的激活情况, 如果 j 是一个卷积层, 那么 $\phi_j(x)$ 将是形状 $C_j \times H_j \times W_j$ 的特征映射, 特征重建损失是特征表示之间的欧式距离, 如 eq 8。

Fig.1表示经过训练以将输入图像转换为输出图像的神经网络。用于图像分类的预训练损失网络有助于通知损失函数。预先训练的网络有助于定义测量图像之间内容和风格的感知差异所需的感知损失函数。

¹¹<https://deepai.org/machine-learning-glossary-and-terms/perceptual-loss-function>

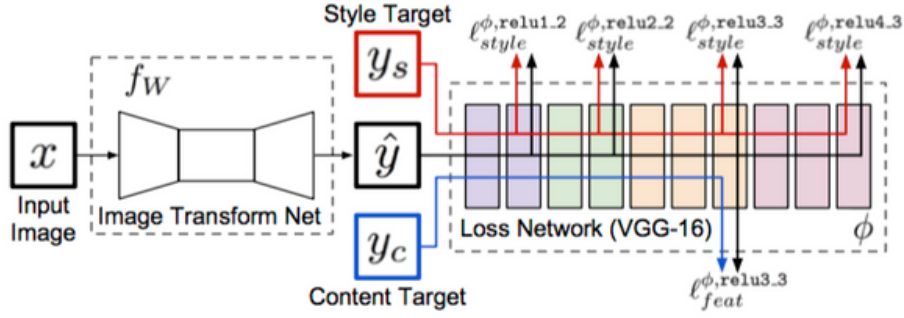


Figure 1: System overview. We train an image transformation network to transform input images into output images. We use a loss network pretrained for image classification to define perceptual loss functions that measure perceptual differences in content and style between images. The loss network remains fixed during the training process.

对于图像数据来说，网络在提取特征的过程中，较浅层通常提取边缘、颜色、亮度等低频信息，而网络较深层则提取一些细节纹理等高频信息，再深一点的网络层则提取一些具有辨别性的关键特征，也就是说，网络层越深提取的特征越抽象越高级。

感知损失就是通过一个固定的网络 (通常使用预训练的 VGG16 或者 VGG19)，分别以真实图像 (Ground Truth)、网络生成结果 (Prediction) 作为其输入，得到对应的输出特征: feature_gt、feature_pre，然后使用 feature_gt 与 feature_pre 构造损失 (通常为 \mathcal{L}_2 -loss)，逼近真实图像与网络生成结果之间的深层信息，也就是感知信息，相比普通的 \mathcal{L}_2 -loss 而言，可以增强输出特征的细节信息¹²。

SSIM loss function

SSIM 损失函数是一种用于衡量两幅图像之间差距的损失函数。它考虑了亮度、对比度和结构指标，这就考虑了人类视觉感知，一般而言，SSIM 得到的结果会比 \mathcal{L}_1 -loss, \mathcal{L}_2 -loss 的结果更有细节¹³。

每个像素 p 的SSIM被定义为

$$\begin{aligned} \text{SSIM}(p) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ &= l(p) \cdot cs(p) \end{aligned} \quad (9)$$

其中省略了均值和标准偏差对像素 p 的依赖性，均值和标准差是用标准偏差为 σ_G, G_{σ_G}

$$\begin{aligned} \varepsilon(p) &= 1 - \text{SSIM}(p) : \\ \mathcal{L}^{\text{SSIM}}(P) &= \frac{1}{N} \sum_{p \in P} 1 - \text{SSIM}(p). \end{aligned} \quad (10)$$

¹²https://blog.csdn.net/qq_43665602/article/details/127077484

¹³<https://blog.csdn.net/u013289254/article/details/99694412>

eq. 9表明 $\text{SSIM}(p)$ 需要关注像素 p 的邻域，这个领域的大小取决于 G_{σ_G} ，网络的卷积性质允许我们将 SSIM 损失写为

$$\mathcal{L}^{\text{SSIM}}(P) = 1 - \text{SSIM}(\tilde{p}). \quad (11)$$

其中 \tilde{p} 是 P 的中心像素。

MS-SSIM loss function

多尺度结构相似性 (MS-SSIM) 损失函数是基于多层 (图片按照一定规则，由大到小缩放) 的 SSIM 损失函数，相当于考虑了分辨率¹⁴。它是一种更为复杂的 SSIM 损失函数，可以更好地衡量图像之间的相似性。

$$\text{MS-SSIM}(p) = l_M^\alpha(p) \cdot \prod_{j=1}^M cs_j^{\beta_j}(p) \quad (12)$$

其中 M, j 描述的是比例，设 $\alpha = \beta_j = 1$ ，对于 $j = 1, \dots, M$ 类似 eq. 11，利用中心像素 \tilde{p} 处计算的损失来近似贴片 P 的损失：

$$\mathcal{L}^{\text{MS-SSIM}}(P) = 1 - \text{MS-SSIM}(\tilde{p}) \quad (13)$$

Cross-entropy loss function

交叉熵损失函数是一种常用的分类问题损失函数。在二分类问题中，它的定义为 Eq. 14

$$\mathcal{L}(\hat{y}, y) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})) \quad (14)$$

其中， \hat{y} 表示模型预测的概率值 (即分类器输出)， y 表示样本真实的类别标签。对于正例样本 ($y = 1$)，交叉熵损失函数的值等于 $\log \hat{y}$ ；对于反例样本 ($y = 0$)，交叉熵损失函数的值等于 $\log(1 - \hat{y})$ 。因此，交叉熵损失函数的目标是最小化模型预测与实际标签之间的差距，从而让模型能够更准确地进行分类。

交叉熵损失函数可以推广到多分类问题中，此时它的表达式略有不同。在多分类问题中，交叉熵损失函数可以写成以下形式：

$$\mathcal{L}(\hat{y}, y) = - \sum_{i=1}^K y_i \log \hat{y}_i \quad (15)$$

¹⁴<https://blog.csdn.net/u013289254/article/details/99694412>

其中, K 表示类别的数量, y_i 表示第 i 个类别的真实标签, \hat{y}_i 表示模型对于第 i 个类别的预测概率值。交叉熵损失函数的目标仍然是最小化预测与实际标签之间的差距, 从而让模型能够更准确地进行分类。

Adversarial loss function

Adversarial loss function 是生成对抗网络 (GAN) 的标准损失函数。GAN 是由生成器 (Generator) 和判别器 (Discriminator) 组成的两个神经网络模型。生成器的目标是生成与真实数据相似的假数据, 而判别器的目标是将真实数据与生成的假数据区分开来。Adversarial loss 使得 GAN 能够不断的改进生成器的性能, 使其能够生成更加逼真的数据。

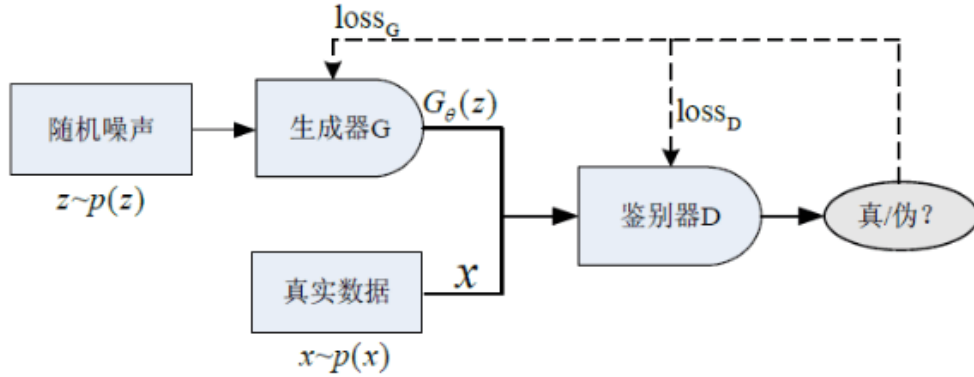


Figure 2: Computational flow and structure of the GAN.

如 Fig. 2所示, 在训练过程中, 生成器和判别器相互竞争和对抗。生成器试图生成逼真的假数据 $G_\theta(z)$ 以欺骗判别器, 而判别器则试图准确地判断真实数据和生成的假数据。Adversarial loss 通过衡量生成器生成的假数据被判别器识别为真实数据的程度, 来指导生成器的训练。它的目标是最小化生成器生成的假数据与真实数据之间的差异, 使得判别器难以区分它们。

一般而言, GAN 的目标函数为 $V(D, G)$ Eq. 17

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_{noise}(z)} [\log (1 - D(G(z)))] \quad (16)$$

其中, $E(*)$ 表示分布函数的期望值, $P_{data}(x)$ 代表着真实样本的分布, $P_{noise}(z)$ 是定义在低维的噪声分布, 通过参数为 θ_g 的 G 映射到高维的数据空间得到 $P_g = G(z, \theta_g)$, 这些参数通过交替迭代的方法来进行优化, 见 Fig. 3。Fig. 3a固定 G 参数不变, 优化 D 的参数, 即最大化 $\max V(D, G)$ 等价于 $\min [-V(D, G)]$ 。因此, D 的损失函数等价于 Eq. 16

$$J^D(\theta^D, \theta^G) = -\mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] - \mathbb{E}_{\tilde{x} \sim p_g(x)} [\log (1 - D(\tilde{x}))] \quad (17)$$

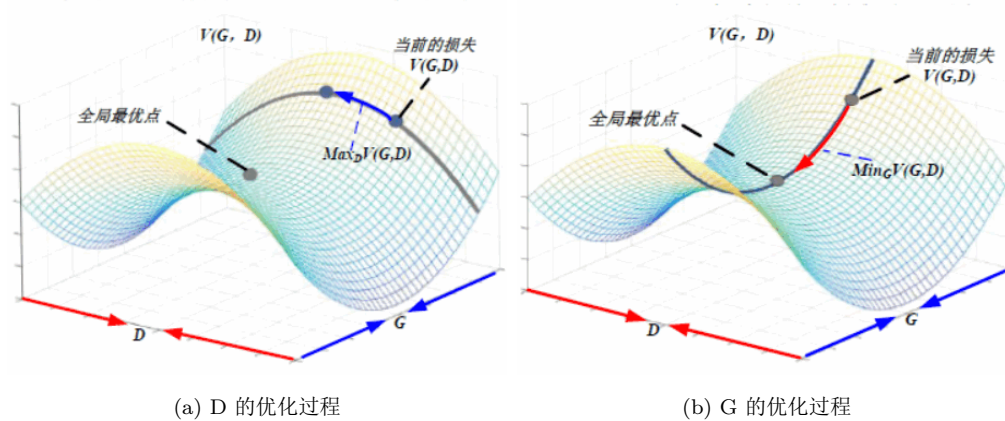


Figure 3: The optimization of the GAN parameters

Adversarial loss 的具体形式可以根据具体的 GAN 架构和任务而定，常见的形式包括最小二乘损失 (Least Squares Loss)、二进制交叉熵损失 (Binary Cross-Entropy Loss) 等。这些损失函数的选择取决于具体的生成器和判别器结构以及任务的特点¹⁵。

Region loss function

Region loss function 是一种用于生物医学图像分割的损失函数。它是一种多功能的损失函数，可以同时考虑类别不平衡和像素重要性，并且可以很容易地实现为 softmax 输出和 RW(Region-wise) 地图之间的像素级乘法¹⁶。一般在一些涉及到异常检测的图像处理中，可以通过设计 region loss function 来使网络能够学习异常区域的位置，但是一般在这种情况下，需要设计一种可训练的异常区域引导框架。

在暗光增强领域，由于难以同时处理包括亮度、对比度、伪影和噪声在内的各种因素，该问题具有挑战性，在损失函数部分，结合使用 Region loss function 能获得不错的效果。一种基于深度学习的微光图像增强方法 (MBLLEN) 的损失函数的详细信息如 Fig. 4所示，Structural loss 损失旨在改善输出图像的视觉质量，Context loss 则通过关注图像中更加高层的信息来提高图像的视觉质量。

上述损失函数将图像作为一个整体。然而，对于弱光增强任务，我们需要更多地关注那些弱光区域。因此，作者提出了 Region loss 损失，它平衡了弱光和图像中其他区域的增强程度。为了做到这一点，作者首先提出了一个简单的策略来分离图像中的弱光区域。通过初步的实验，作者发现在所有像素中选择最暗的 40% 的像素可以很好地近似于弱光区域。人们也可以提出更复杂的方法来选择暗区，事实上在文献中有很多。最后，区域损失定义

¹⁵<https://blog.csdn.net/qikaihuting/article/details/84950947>

¹⁶<https://arxiv.org/abs/2108.01405>

如 Eq. 18

$$L_{Region} = w_L \cdot \frac{1}{m_L n_L} \sum_{i=1}^{n_L} \sum_{j=1}^{m_L} (\|E_L(i, j) - G_L(i, j)\|) + w_H \cdot \frac{1}{m_H n_H} \sum_{i=1}^{n_H} \sum_{j=1}^{m_H} (\|E_H(i, j) - G_H(i, j)\|) \quad (18)$$

其中, E_L 和 G_L 是增强图像和地面实况的暗光区域, E_H 和 G_H 是图像的剩余部分。

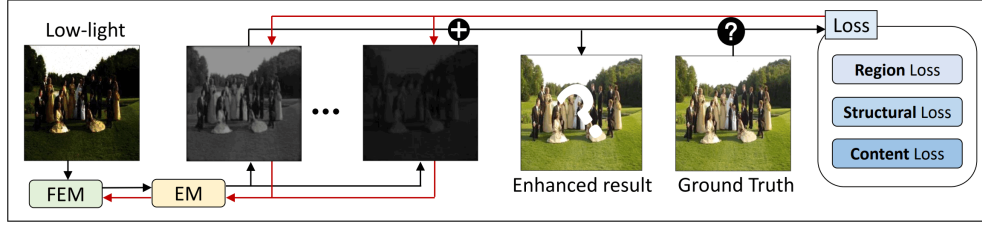


Figure 4: Data flow for training. The proposed loss function consists of three parts.

Reflectance loss function

Reflectance loss function 是计算机视觉领域中的一种损失函数,常用于图像去雾 (image dehazing) 任务中。(图像去雾是指通过算法去除图像中由雾霾引起的能见度降低的效果。Reflectance loss function 用于衡量生成的去雾图像与真实清晰图像之间的差异,以指导去雾模型的训练过程。)

Reflectance loss function 的核心思想是基于图像的反射率 (reflectance) 属性。反射率是指物体表面对入射光线的反射程度,与雾霾相关的信息主要存在于反射率中。通过计算生成的去雾图像的反射率与真实清晰图像的反射率之间的差异,可以量化生成图像的质量。

具体来说, Reflectance loss function 通常使用像素级别的差异度量函数,比如均方误差 (Mean Squared Error, MSE) 或结构相似性 (Structural Similarity, SSIM) 等来计算生成图像与真实图像之间的差异。优化过程的目标是最小化 Reflectance loss,使得生成的去雾图像能够与真实清晰图像更加接近。

通过引入 Reflectance loss 作为训练目标,去雾模型可以学习到更准确的雾霾分布和反射率信息,从而生成更清晰的去雾图像。这有助于提高图像去雾算法的性能和质量。

Consistency loss function

Consistency loss function 是一种在机器学习中使用的损失函数,用于提高模型的一致性。它主要用于半监督学习或自监督学习任务中,其中存在一些带有标签的数据(有目标变量的输入-输出对)和一些没有标签的数据(只有输入)。区别于全监督学习,半监督学习针对训练集标记不完整的情况:仅仅部分数据具有标签,然而大量数据是没有标签的。因

此，目前半监督学习的关键问题在于如何充分地挖掘没有标签数据的价值。主流的半监督学习方法有下面几种¹⁷：

(1) **自训练方法 (Self-Training)**。这是一种很直观的思路：既然大量数据是没有标签的，那么能否对这些数据生成一些伪标签 (Pseudo Labels)，对这些伪标签数据的训练从而利用原始的无标签数据。

(2) **基于对抗学习的方法 (Adversarial-Learning-based)**。这类方法基于一种假设，无标签数据 (unlabeled data) 通常具有和有标签数据 (labeled data) 在某种程度上类似的潜在标签。所以很自然地，可以采用 GAN 的图像模拟思路来进行对有标签数据进行类似于无标签数据的数据增强，进而利用无标签数据的潜在知识 (latent knowledge)。

(3) **基于一致性的方法 (Consistency-based)**。这类方法的核心思路在于一致性损失函数 Consistency loss function，对于经过扰动的无标签数据，模型应该对其做出一致性的预测——可以理解成一种利用无标签数据进行网络正则化的方法。

一致性损失函数的基本思想是，在模型对相似输入的处理应该具有一致性。它通过对输入进行一些变换或扰动，然后要求模型在这些变换后仍然能够产生相似的输出。如果模型能够在不同的变换下产生一致的输出，那么我们可以认为模型具有一定的鲁棒性和泛化能力。

具体来说，一致性损失函数可以定义为模型在原始输入和变换后输入上的输出之间的差异。常见的一致性损失函数包括均方差损失、平均绝对误差损失、KL 散度等。通过最小化一致性损失函数，模型被迫保持对输入的一致性响应，从而提高模型的泛化能力。

需要注意的是，一致性损失函数的具体形式和应用取决于具体的任务和模型架构。它通常与其他损失函数（如分类损失函数或重构损失函数）结合使用，以共同训练模型。

Color loss function

$$\mathcal{L}_c^i = \sum_p \angle \left((\mathcal{F}(I_i))_p, (\tilde{I}_i)_p \right) \quad (19)$$

Lim [4] 基于余弦相似性基于 [5] 中提出的损失函数 color-loss 进行了改进，针对于颜色恢复设计的损失函数，将 RGB 三通道看做三维向量，计算增强图像与 GT 之间的余弦距离 (见 Eq. 19)，因为在 reconstruction loss 中使用了 \mathcal{L}_2 -loss，但是 \mathcal{L}_2 -loss 只是数值上测量色差，不能保证颜色向量的方向是相同的，可能导致颜色 mismatch，所以使用了夹角的 loss，Lim [4] 修改之后，见 Eq. 20。

$$\mathcal{L}_c = \sum_{k=1}^3 \left(1 - \frac{1}{H_k \times W_k} \sum_{i \in (H_k, W_k)} \frac{\tilde{I}_k(i) \cdot I_k^*(i)}{\|\tilde{I}_k(i)\|^2 \|I_k^*(i)\|^2} \right) \quad (20)$$

¹⁷https://blog.csdn.net/JYZhang_CVML/article/details/106817709

其中 H_k 和 W_k 分别是第 k 个拉普拉斯金字塔级别 (Laplacian pyramid level) 中的增强结果的高度和宽度。 $\tilde{I}_k(i)$ 表示像素位置, \cdot 表示内积。颜色越接近, 颜色损失 L_c 就越接近于零。通过在恢复过程中考虑颜色向量的方向, 能够成功地将增强结果的颜色属性恢复为与地面实况相似。

文章描述该方法是通过模糊输入图像与 GT 的纹理、内容, 仅仅保存图像的颜色信息实现图像颜色的校正。实现过程比较简单, 首先构建一个高斯模糊核, 然后利用高斯模糊核作为卷积核对图像进行卷积运算, 得到模糊后的图像; 然后计算输入图像与 GT 的 MSE 作为损失函数。(pytorch 实现¹⁸)

Laplacian loss function

拉普拉斯损失 (Laplacian loss) 是一种在计算机视觉任务中常用的损失函数, 用于促进图像的平滑性。它通常用于图像生成、图像修复或图像去噪等任务中。其形式上一般定义为 Eq. 21。从形式上看, Laplacian loss function 主要工作是在 \mathcal{L}_1 -loss 的基础上乘了一个 2 的幂次项, 并且将多层的结果相加¹⁹。

Laplacian loss 的基本思想是鼓励生成的图像具有平滑的变化, 以减少图像中的噪点或锐利边缘。它利用图像的拉普拉斯金字塔来比较生成图像和真实图像之间的结构差异。拉普拉斯金字塔 [6] 最初是为紧凑的图像表示而设计的, 具有简单的操作, 即缩小和扩展。许多研究人员试图将拉普拉斯金字塔应用于广泛的计算机视觉任务, 例如纹理分析和合成。拉普拉斯金字塔具有线性可逆过程来重建原始输入而不损失任何信息以精确地细化局部细节的性质。Ghiasi 等人 [7] 提出了基于拉普拉斯金字塔的细化过程, 以增强语义分割的低分辨率估计结果的局部细节。

$$\mathcal{L}_l = \sum 2^{i-1} \|\mathcal{L}^i(\hat{\alpha}) - \mathcal{L}^i(\alpha)\|_1 \quad (21)$$

具体来说, Laplacian loss 首先通过将生成图像和真实图像分别构建成拉普拉斯金字塔。拉普拉斯金字塔是一种多尺度表示, 其中每个级别表示了原始图像与其上一级别平滑版本之间的细节差异。然后, Laplacian loss 计算两个图像金字塔之间的差异, 通过比较它们在每个级别上的像素差异来量化结构上的差异。

通过最小化 Laplacian loss, 生成的图像被鼓励具有更平滑的结构, 从而减少图像中的噪点和锐利边缘。

需要注意的是, 具体的 Laplacian loss 的实现方式可能因任务而异。在某些情况下, 可以使用像素级别的差异或特征级别的差异作为损失度量。此外, Laplacian loss 通常与其他损失函数 (如像素级别的差异、感知损失或对抗损失) 结合使用, 以共同训练生成模型。

¹⁸<https://blog.csdn.net/ztzi321/article/details/101424448>

¹⁹<https://blog.csdn.net/lanceloter/article/details/121925904>

1.2 Paper reading

1.2.1 Switching gaussian mixture variational rnn for anomaly detection of diverse cdn websites [1]

Known problems

问题一：单个网站在不同时间段的非平稳依赖（The Non-stationary Dependencies）会降低深度异常检测模型的性能：

由于用户的正常行为或 CDN 的调度等，相应的 KPI 通常表现出非平稳的时间特征，不应将其归类为服务失败或退化。但是，这些类型的预期模式很难被目前的方法捕获，这会使得目前的模型在 CDN KPI 异常检测方面的表现性能很差。

从 Fig. 5a和 Fig. 5d可以明显看出，工作日和周末的用户请求行为是不同的。前一个网站显示了在周末相对于工作日的用户请求激增，而后一个数据展示了相反的情况。

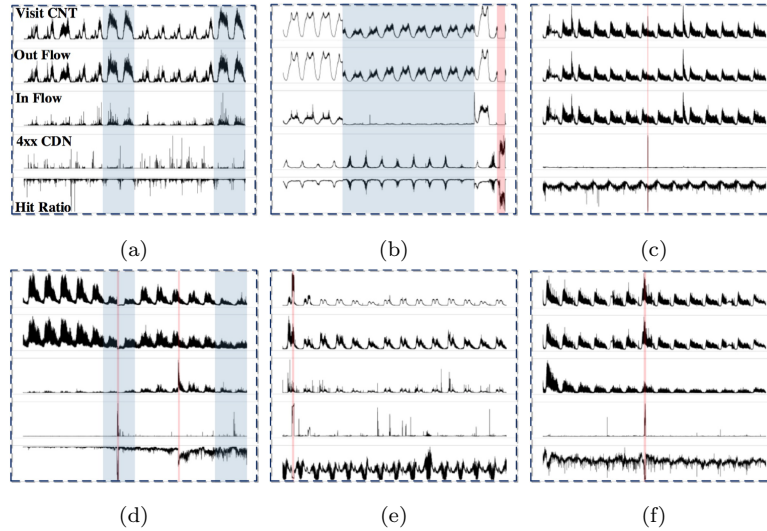


Figure 5: 2-weeks real world typical multivariate CDN KPIs of 6-websites. Periods in light blue show the change points in KPIs; Regions highlighted

问题二：不同的网站在 CDN KPI 中表现出不同的特征，但其中又有一些相似特征。这使得现有的单个深度异常检测模型无法良好的捕捉这种动态复杂性。

Fig. 5b是另一种典型情况，即部分用户被 CDN 的调度中心调度到另一组边缘节点上；所以，KPI 的变化时刻，发生在这段时间内。Fig. 5a中的视频点播网站的 KPI 通常与 Fig. 5e中的直播网站有很大的不同。

因为一般商业 CDN 会为数百甚至数千个网站提供服务，这些网站会因为服务类型和用户的请求行为的不同而表现出不同的特征，因此可以观察到不同网站的 KPI 在空间和时间特征上的变化。

问题三：资源浪费

为了对多个网站进行有效的异常检测，现有的深度异常检测方法通常为每个网站训练一个单独的模型，从而存在为每个网站训练和维护大量的单独模型的问题，不仅消耗巨大的计算和存储资源，但也增加了模型维护的成本。

例如，Fig. 5c和 Fig. 5f显示出相似的特征，但它们属于不同网站的 KPI。在这种情况下，为每个网站训练一个单独的模型完全是浪费。

Challenge

现有的许多深度学习方法，模型学习正常情况下的网络 KPI，以构建并训练模型，从而应用于无监督的网络 KPI 异常检测，即深度异常检测。

CDN KPI 数据存在的问题:

(1) CDN 下具有多个节点，每个节点网络 KPI 曲线的多样性，即 KPI 曲线表现为周期型的，有稳定型的，也有不稳定型的。

(2) CDN 下往往具备大量的节点网络，这些网络的异常种类多，核心网网元数据多，故障发生的类型也多种多样，导致了异常种类也多种多样，即 KPI 之间呈现出非平稳的顺序关系。

Author Work

(1) 表征不同的复杂 KPI 时间序列结构和动态特征；

(2) 基于以上问题，现有的深度学习方法非常难以表示和识别 CDN 网下节点网络的异常，因此作者提出了一种方法，一种适用于多变量 CDN KPI 的切换高斯混合变分循环神经网络 (SGmVRN)。

1. SGmVRNN 引入变分循环结构，并将其潜在变量分配到混合高斯分布中，以对复杂的 KPI 时间序列进行建模并捕捉其中的不同结构和动态特征，而在下一步中，它结合了一种切换机制来表征这些多样性，从而学习更丰富的 KPI 表示。

2. 为了有效的推理，我们开发了一种向上向下的自动编码推理方法，它结合了参数的自下而上的似然性和自下而上的先验信息，以实现准确的后验近似。

(3) 结果：通过 SGmVRNN 方法构造出来的模型应用到不同网站的 CDN KPI 异常识别，SGmVRNN 模型的平衡 F 分数明显高于目前最优秀的模型。

Proposed Method

SGmVRNN: 将概率混合和切换机制融合到 VRNN 中，它将概率混合和切换机制融合到 VRNN 中，从而有效地模拟单个网站的多元 CDN KPI 相邻时间步之间的非平稳时间依

赖性，以及不同网站之间的动态特性。

SGmVRNN 特点：能够在不同的时间学习不同的结构特征，并捕捉它们之间的各种时间依赖性，以表征不同 CDN 网站的多元 KPI 中的复杂结构和动态特征。

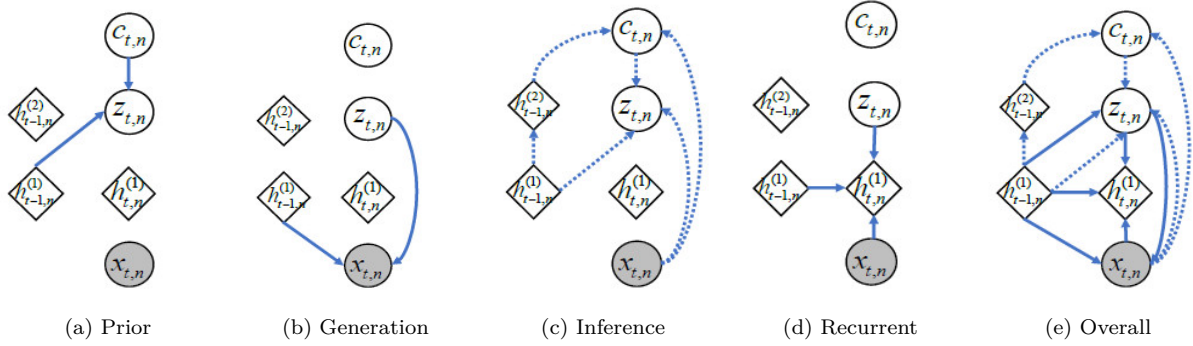


Figure 6: Graphical illustration of each operation of the SGmVRNN: (a) conditional prior of latent variables $z_{t,n}$ and $c_{t,n}$; (b) generation process of $x_{t,n}$; (c) inference of the variational distribution of $z_{t,n}$ and $c_{t,n}$; (d) updating the hidden units of the RNN recurrently; (e) overall operations of the SGmVRNN. Note that circles denote stochastic variables while diamond-shaped units are used for deterministic variables, and shaded nodes denote observed variables.

如何解决主要挑战？

将切换机制 (见 Fig. 6) 与混合模型相结合，SGmVRNN 可以利用切换机制对电流输入的刻画和多种时间变化的传输提供足够的表示能力，从而解决第一个问题。

利用混合高斯分布潜变量很好地处理多样化的网站，从而解决第二个问题。

2 个人工作进展

2.1 关于“华为杯”的一点思考

2.1.1 适用多场景的通用化时序预测算法

[1] 中提出的方法 SGmVRNN 在可以表征复杂数据，面对拥有序列多样性的数据集 (题目中提到数据可能具有不同长度、趋势性、周期性、采样粒度)，作者提出了可以把潜在变量 (latent variable) 分配到混合高斯分布中，从而可以对复杂的时间序列进行建模。切换机制似乎可以较好的满足通用化这样一种需求。

作者引入切换机制是为了解决 CDN 网络下非平稳的时间依赖性以及不同网站的 KPI 特征，题目描述中针对的是云计算的场景，如资源的消耗序列、用户的行为序列、机房传感器的读数序列，对不同的场景设计不同的算法，显然不太现实。二者可能具有类通性。

2.2 梳理损失函数

这一周的工作主要是接着梳理上一周末梳理完成的损失函数

1) **Consistency loss function** 一般用于半监督学习中，用于对 unlabeled data 进行变换和扰动，如果模型在这些不同的变换下，结果相似或者一致，就可以说明模型具有一定的鲁棒性和泛化性。同样，其没有具体的表现形式，取决于具体的任务和模型架构，一般情况下不单独使用，而是与其他损失函数结合使用，共同训练模型。此外总结了三种主流的半监督学习方法。

2) **Color loss function** 是一种实现图像颜色的校正的损失函数。一般结合其他损失函数使用，在结合 \mathcal{L}_2 -loss 时，因为 \mathcal{L}_2 -loss 只是能够从数值上测量色差，无法保证颜色向量的方向是相同的，这可能会导致颜色 mismatch，Color loss 需要计算夹角余弦。

3) **Laplacian loss function** 则是衡量预测结果与 GT 之间对应像素上的颜色差异。使用这个损失能够得到更好的结果。它还能捕捉全局以及局部的差异。此外，额外去了解了一下拉普拉斯金字塔的详细原理。

从数学角度看，图像的高斯模糊过程就是图像与正态分布做卷积，在二维卷积里面，一个正方形高斯核的值中间高、四周边缘低，大致呈一个“圆形”，拉普拉斯金字塔与高斯金字塔相反，是个倒着的金字塔，也就是图片尺度逐级变大，但是这个并不是单纯把图像放大，它要计算 [原图像] 和 [下采样、上采样后图像] 两者差值，得到的结果尺度逐渐变大，形成一个金字塔。高斯金字塔下采样过程中丢失的信息在上采样后并不能完全恢复，丢失信息的情况依然存在，如果直观把原图和下采样、上采样后的图片进行对比，会发现后者仍然模糊，此时用前者减去后者得到的新图像，形成的就是拉普拉斯金字塔。这个相减的过程，能够更突出显示图像的像素信息，因为高斯核属于低通滤波，图像经过这个操作后得到的结果更加平滑，边界信息相比原图没那么明显（因为图像中高频的边缘部分都被过滤），而原图这些边界信息较为明显，至于非边界部分信号变化不大，因此相减后得到的图像更关注边界的信息，或者说对高频信号更为敏感。

4) 阅读文献《Switching gaussian mixture variational rnn for anomaly detection of diverse cdn websites》，作者提出了一种适用于多元 CDN kpi 的开关高斯混合变分递归神经网络 (SGmVRNN)。具体来说，SGmVRNN 引入了变分循环结构和分配其潜在变量到混合高斯分布模型复杂 KPI 时间序列和捕获各种结构和动态特征，而在下一步它包含一个转换机制来描述这些多样性，从而学习丰富的 KPI 表示。

为了进行有效的推理，作者提出了一种向上-向下的自编码推理方法，该方法结合了自下而上的似然值和参数的上下先验信息，以实现精确的后验逼近。在真实世界数据上进行的大量实验表明，根据来自不同网站的 CDN kpi 上的 f1 分数，SGmVRNN 的性能显著优于最先进的方法。

2.3 复现 KinD 代码

KinD²⁰采用的是类 U-Net 网络，且对比使用了多种损失函数，采用的数据集相对较少，且环境配置相对简单，具体见 Tab. ??。同时，KinD [8] 目前引用数量为 543，采用的网络架构较为经典，即 Retinex-based 的思想。

2.3.1 Requirement

```
1 Python
2 Tensorflow >=1.10.0
3 numpy, PIL
4
```

2.3.2 Model

lrelu

```
1 def lrelu(x, trainbale=None):
2     return tf.maximum(x*0.2,x)
3
```

`lrelu` 函数实现了带有泄露的线性整流单元 (Leaky ReLU)，其作用是在神经网络中引入非线性变换。

线性整流单元 (ReLU) 是一种常用的激活函数，对于输入 x ，当 x 大于等于零时，输出为 x ，当 x 小于零时，输出为零。这种函数在训练神经网络时具有很好的性质，但是在某些情况下，当输入为负值时，导数为零会导致梯度消失的问题。

为了解决这个问题，Leaky ReLU 引入了一个小的负斜率，在输入为负值时，输出为输入乘以一个小的斜率。公式表示如下：

$$lrelu(x) = \max(x * \alpha, x)$$

input

其中， x 是输入， α 是小的斜率（一般取很小的正数，如 0.2），`max` 函数选择两者中的较大值作为输出。

²⁰<https://github.com/zhangyhuace/KinD>

function

Leaky ReLU 通过引入负斜率，保留了负值的信息，从而解决了梯度消失问题，并且具有更好的收敛性和表达能力。

在给定的代码中，lrelu 函数接受输入 x 和一个可选的参数 trainable（默认为 None），并返回 x 乘以斜率 0.2 和 x 中的较大值作为输出。该函数可以作为激活函数在神经网络的层中使用，引入非线性变换，增加网络的表达能力。

upsample_and_concat

```
1      def upsample_and_concat(x1, x2, output_channels, in_channels, scope_name,
                                trainable=True):
2      with tf.variable_scope(scope_name, reuse=tf.AUTO_REUSE) as scope:
3          pool_size = 2
4          deconv_filter = tf.get_variable('weights', [pool_size, pool_size,
                                                         output_channels, in_channels],
                                                         trainable= True)
5          deconv = tf.nn.conv2d_transpose(x1, deconv_filter, tf.shape(x2) , strides=[1,
                                                         pool_size, pool_size, 1], name=
                                                         scope_name)
6
7          deconv_output = tf.concat([deconv, x2], 3)
8          deconv_output.set_shape([None, None, None, output_channels*2])
9
10     return deconv_output
11
```

函数的作用是将两个特征图进行上采样和拼接操作。在图像处理中，上采样是指将特征图的尺寸扩大，通常使用插值等方法来实现。而拼接操作则是将两个特征图在通道维度进行连接。

这个函数常用于 U-Net 等图像分割和生成网络中，用于实现特征图的跳跃连接，从而传递低层次的细节信息到高层次的特征图中，提升模型性能和图像质量。

input

x1: 第一个输入特征图。

x2: 第二个输入特征图。

output_channels: 输出特征图的通道数。

in_channels: 输入特征图的通道数。

scope_name: 变量作用域的名称。

trainable: 是否可训练的布尔值, 默认为 True。

function

在给定的变量作用域下, 创建一个名为'weights'的变量, 它是一个大小为 [pool_size, pool_size, output_channels, in_channels] 的可训练的卷积核。

使用 tf.nn.conv2d_transpose 函数对 x1 进行转置卷积操作 (也称为反卷积), 使用创建的卷积核 deconv_filter, 步长为 [1, pool_size, pool_size, 1], 以获得上采样后的特征图 deconv。

将 deconv 和 x2 进行通道维度上的拼接, 通过 tf.concat 函数将它们在通道维度上连接起来, 得到 deconv_output。

设置 deconv_output 的形状为 [None, None, None, output_channels*2], 其中 None 表示该维度的大小可以根据输入的 shape 自动推断。

返回拼接后的特征图 deconv_output。

DecomNet_simple

```
1  def DecomNet_simple(input):
2      with tf.variable_scope('DecomNet', reuse=tf.AUTO_REUSE):
3          conv1=slim.conv2d(input,32,[3,3], rate=1, activation_fn=lrelu,scope='
                                     g_conv1_1')
4          pool1=slim.max_pool2d(conv1, [2, 2], stride =2, padding='SAME' )
5          conv2=slim.conv2d(pool1,64,[3,3], rate=1, activation_fn=lrelu,scope='
                                     g_conv2_1')
6          pool2=slim.max_pool2d(conv2, [2, 2], stride =2, padding='SAME' )
7          conv3=slim.conv2d(pool2,128,[3,3], rate=1, activation_fn=lrelu,scope='
                                     g_conv3_1')
8          up8 = upsample_and_concat( conv3, conv2, 64, 128 , 'g_up_1')
9          conv8=slim.conv2d(up8, 64,[3,3], rate=1, activation_fn=lrelu,scope='
                                     g_conv8_1')
10         up9 = upsample_and_concat( conv8, conv1, 32, 64 , 'g_up_2')
11         conv9=slim.conv2d(up9, 32,[3,3], rate=1, activation_fn=lrelu,scope='
                                     g_conv9_1')
12         # Here, we use 1*1 kernel to replace the 3*3 ones in the paper to get
                                     better results.
13         conv10=slim.conv2d(conv9,3,[1,1], rate=1, activation_fn=None, scope='
                                     g_conv10')
14         R_out =tf.sigmoid(conv10)
```

```

15
16     l_conv2=slim.conv2d(conv1,32,[3,3], rate=1, activation_fn=lrelu,scope='
                                l_conv1_2')
17     l_conv3=tf.concat([l_conv2, conv9],3)
18     # Here, we use 1*1 kernel to replace the 3*3 ones in the paper to get
                                better results.
19     l_conv4=slim.conv2d(l_conv3,1,[1,1], rate=1, activation_fn=None,scope='
                                l_conv1_4')
20     L_out =tf.sigmoid(l_conv4)
21
22     return R_out, L_out
23

```

函数是一个图像分解网络 (Image Decomposition Network) 的简化版本。它的主要作用是将输入的图像分解为反射分量 (Reflection component) 和照明分量 (Illumination component) 两部分。

函数接受一个输入图像作为参数，并在 TensorFlow 的变量作用域'DecomNet' 下构建网络。该网络包含一系列的卷积、池化和上采样操作，以及一些连接操作和卷积层。

input

input (一个张量，作为输入图像)。

function

实现图像的分解网络。

具体地，DecomNet_simple 函数的网络结构如下：

1. 输入图像经过第一层卷积层 conv1，输出通道数为 32，卷积核大小为 3x3，使用 Leaky ReLU 激活函数。
2. 经过池化层 pool1，进行最大池化，池化窗口大小为 2x2，步长为 2。
3. 经过第二层卷积层 conv2，输出通道数为 64，卷积核大小为 3x3，使用 Leaky ReLU 激活函数。
4. 经过池化层 pool2，进行最大池化，池化窗口大小为 2x2，步长为 2。
5. 经过第三层卷积层 conv3，输出通道数为 128，卷积核大小为 3x3，使用 Leaky ReLU 激活函数。
6. 调用 upsample_and_concat 函数，将 conv3 和 conv2 进行上采样并连接起来，输出通道数为 64。

7. 经过第八层卷积层 conv8, 输出通道数为 64, 卷积核大小为 3x3, 使用 Leaky ReLU 激活函数。

8. 调用 upsample_and_concat 函数, 将 conv8 和 conv1 进行上采样并连接起来, 输出通道数为 32。

9. 经过第九层卷积层 conv9, 输出通道数为 32, 卷积核大小为 3x3, 使用 Leaky ReLU 激活函数。

10. 经过最后一层卷积层 conv10, 输出通道数为 3, 卷积核大小为 1x1, 不使用激活函数。该层用于生成反射分量 (Reflection component)。

11. 使用 sigmoid 函数对 conv10 的输出进行映射, 得到反射分量 R_out。

12. 经过第一层卷积层 l_conv2, 输出通道数为 32, 卷积核大小为 3x3, 使用 Leaky ReLU 激活函数。

13. 将 l_conv2 和 conv9 进行连接操作, 得到连接特征图 l_conv3。

14. 经过一层卷积层 l_conv4, 输出通道数为 1, 卷积核大小为 1x1, 不使用激活函数。该层用于生成照明分量 (Illumination component)。

15. 使用 sigmoid 函数对 l_conv4 的输出进行映射, 得到照明分量 L_out。

最终, DecomNet_simple 函数返回反射分量 R_out 和照明分量 L_out。这个函数的目标是将输入图像分解为反射分量和照明分量, 以便进一步进行图像修复或图像增强等后续处理。

3 下周工作计划

(1) 继续复现、分析 KinD 的项目代码, 目前架构中的三个结构, 从 model.py 中读取到了 5 个函数 lrelu、upsample_and_concat、DecomNet_simple、Restoration_net、Illumination_adjust_net, 目前主要看了前三个函数, 文档中有对应说明, 下周计划把后面两个函数补完, 看看能不能加上图, 同时梳理下每一层的功能。

(2) 想要去了解一下“为什么高斯核用于图像模糊, 拉普拉斯核用于图像锐化”的数学原理。

References

- [1] Liang Dai, Wenchao Chen, Yanwei Liu, Antonios Argyriou, Chang Liu, Tao Lin, Penghui Wang, Zhen Xu, and Bo Chen. Switching gaussian mixture variational rnn for anomaly detection of diverse cdn websites. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pages 300–309. IEEE, 2022.

- [2] Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox. Multi-view 3d models from single images with a convolutional network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, pages 322–337. Springer, 2016.
- [3] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016.
- [4] Seokjae Lim and Wonjun Kim. Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, 23:4272–4284, 2021.
- [5] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6842–6850, 2019.
- [6] Peter J Burt and Edward H Adelson. The laplacian pyramid as a compact image code. In *Readings in computer vision*, pages 671–679. Elsevier, 1987.
- [7] Golnaz Ghiasi and Charless C Fowlkes. Laplacian pyramid reconstruction and refinement for semantic segmentation. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III 14*, pages 519–534. Springer, 2016.
- [8] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia, MM ’19*, page 1632–1640, New York, NY, USA, 2019. Association for Computing Machinery.