

Non-I.I.D. Image Classification

张翔 李家昊 宋浩轩 李浩源

2020 年 5 月 31 日

目录

1 Basic Idea	1
2 Detailed Algorithm	1
2.1 Approach	1
2.2 Implementation Details	2
3 Result	3
3.1 Comparison with Baseline	3
3.2 Ablation Study	3
3.3 Stability	3
4 Collaboration & Contribution	4

1 Basic Idea

对于 NICO Dataset[3], 我们认为它的 Non-IID 特性主要是来源于不同 Context 带来的分布差异。数据集有 10 类动物, 每类动物均会出现在 10 种 Context 中。考虑人类识别场景中动物的过程, 人眼可以将注意力集中在场景中的动物上, 识别出动物的轮廓, 并根据动物的特征判断其所属的类别。理想情况下, 如果模仿人眼识别动物的过程, 能够对图片中的动物进行 Instance Segmentation[4], 此时得到的特征仅与图片的亮度、颜色等相关, 从而能够更好地反映出被分类动物的特点。

在实际实现时, 由于难以得到精细的 Instance mask, 且 Instance Segmentation 对于分类问题来说是一个相对代价较高的操作, 可以考虑让网络能够自动地着重关注图片中动物的本体部分, 即 Attention 机制 [5]。对于输入图片 I , 使用 Attention 模块 θ 生成 mask W_θ , 再将 $W_\theta \otimes I$ 输入到后端的分类器进行训练。当 θ 参数较为合适时, W_θ 近似为需要关注的主体部分, 经过 mask 操作后的图片更容易提取出 context-invariant 的特征。

当前端的特征提取网络固定时, 我们仍然可以对特征向量使用 Attention 机制, 尽量滤除背景 context 的特征, 而保留前景中动物的特征。

2 Detailed Algorithm

2.1 Approach

我们的模型结构如图 1, 主要分为 Residual Attention [6], Classifier, 以及 Discriminator 三个模块。Residual Attention 模块利用注意力机制来强化物体本身的特征, 弱化 context 信息, 其中的 Residual

连接方式有助于网络保持 Identity 映射的能力，类似于 ResNet[2] 的思想；Classifier 用于对物体进行分类；Discriminator 用于判断两张图片是否属于同一 context。

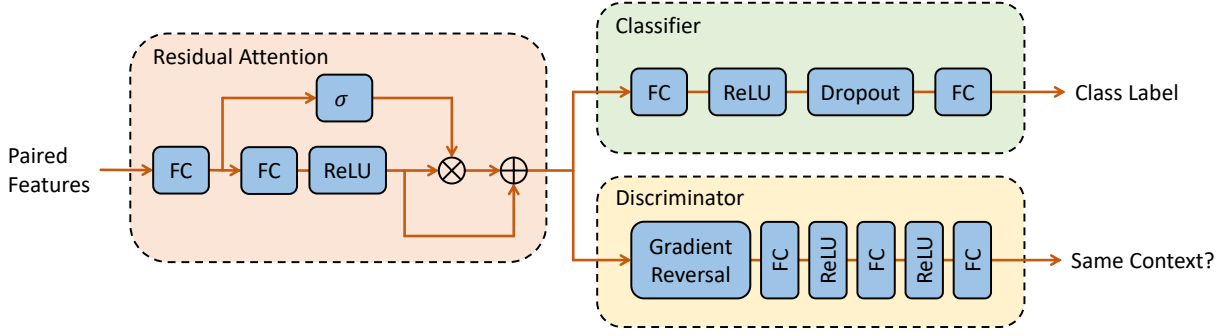


图 1: 模型框架，其中符号 \oplus 表示逐元素相加，符号 \otimes 表示逐元素相乘。

在训练阶段，模型同时接受两张图片的特征作为输入，通过 Residual Attention 优化特征，然后通过 Classifier 和 Discriminator 分支得到分类预测和类别预测。

记模型 Residual Attention, Classifier, 以及 Discriminator 的参数分别为 θ_{att} , θ_{cls} , θ_{dis} , 网络整体的参数为 θ 。传统意义上，我们希望优化分类器和判别器的总体损失函数

$$\min_{\theta} \mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{ctx} \quad (1)$$

其中 \mathcal{L}_{cls} 为 Classifier 分支的 10 分类交叉熵损失， \mathcal{L}_{ctx} 为 Discriminator 分支的 2 分类交叉熵损失，超参 λ 为平衡因子。

考虑到我们需要提取出更加 context-invariant 的特征，对于 \mathcal{L}_{ctx} 构建为如下的 min-max 问题会更加合适

$$\max_{\theta_{att}} \min_{\theta_{ctx}} \mathcal{L}_{ctx} \quad (2)$$

这个式子的意义是 Discriminator 的参数选取将有利于增强它的辨别能力，但 Attention 模块希望通过适当选取特征的 mask，使得提取到的特征能够混淆 Discriminator，即通过 adversarial 的方式来增强双方的能力。

而对于分类器，我们仍然希望增强它的分类能力

$$\min_{\theta_{att}, \theta_{cls}} \mathcal{L}_{cls} \quad (3)$$

为了能够在一次 backward pass 中就能优化所有的参数，这里引入 Gradient Reversal [1] 模块，放在 Discriminator 分支前，它将 $\frac{\partial \mathcal{L}_{ctx}}{\partial \theta_{dis}}$ 反向后传给 Attention 模块，从而实现 min-max 优化。

在测试阶段，我们去掉 Discriminator 模块，模型只接受一张图片的特征作为输入，通过 Residual Attention 模块后，经过 Classifier 预测出最终的类别标签。

2.2 Implementation Details

在给定的训练数据上，每个类共有 7 个 context，我们为每个类划分了 6 个 context 作为训练集，剩余 1 个 context 作为验证集。

在训练阶段，我们选取超参 $\lambda = 1$ ，使用 Adam 优化器训练了 200 个 epoch，batch size 大小为 16，learning rate 为 10^{-5} ，weight decay 为 5×10^{-4} ，选取验证集上准确率最高的模型作为最终模型。

Baseline 我们实现了一些 baseline，其中最简单的是 **Classifier**，它只有两个全连接层；在我们实现的 **AttentionClassifier** 中，可以将 **Discriminator** 更换成一个 **Context Classifier**，它尝试对图片的 **ctx label** 进行分类（实际上这是不够合理的，因为不同 **class** 下的 **context** 是不完全相同的）；其余的 baseline 则在完整版 **AttentionClassifier** 的基础上做 **ablation study** 得到，具体请参考 **3 Result** 部分。

Dataset 在训练时，需要成对输入图片，为此我们实现了一个 **NicoPairedDataset**，它会取同一个 **class** 下相同或不同 **context** 的图片，并根据是否相同赋予 0-1 标签，用以训练 **discriminator**。为了采样的充分与均衡，对每张图片，取同类中不同 **context** 的图片（每个 **context** 随机选取 1 张）与它配对，然后同类相同 **context** 的图片配对，使得这两种配对各占 50%。

3 Result

3.1 Comparison with Baseline

以下测试是基于 6/1 划分训练集与验证集得到的结果。

Name	Accuracy
Classifier	61.82%
Attention Context Classifier	72.30%
Attention Discriminator	74.49%

表 1: Baseline 测试结果

3.2 Ablation Study

测试环境同 Baseline，结果如下

Attention	Residual Attention	Gradient Reversal	Accuracy
✓			73.48%
✓		✓	73.65%
	✓		74.16%
	✓	✓	74.49%

表 2: Ablation Study 测试结果

3.3 Stability

为了测试算法在不同 **context** 下的稳定性和泛化能力，我们将所有的训练样例按照 **context** 分成 4 份。其中 **ctx** 标号为 0-3 的作为训练集，同时以 4/5/6 分别作为 **test1**、**test2** 和 **val** 集。最终我们发现结果在 40 epoch 左右达到稳定，取该 epoch 的结果，测试如下：

Dataset	train 集	val 集	test1	test2
Accuracy	93.94%	66.89%	75.49%	77.47%

表 3: 稳定性测试结果

可见，在只使用 4 个 context 作为训练集的情况下，ctx 为 6 的 val 集并不能达到之前的准确率，但还是有一定提升的。同时，在 val 集达到稳定准确率时，两个 test 集合达到了稳定且很高的准确率，也有比较大的提升。可见，我们的算法在不同的 context 下是比较稳定且具有一定泛化能力的。

4 Collaboration & Contribution

实验分工

- 张翔负责核心算法的设计以及网络的部分实现 (GradientReversal, AttentionResidualBlock, AttentionClassifier 与 NicoPairedDataset)。
- 李家昊负责 Baseline (纯 MLP 网络 Classifier) 的实现以及实验数据处理 (NicoDataset 与 Metrics) 的维护。
- 宋浩轩负责核心算法网络的实现与调整以及稳定性的测试。
- 李浩源负责调节训练参数以及与 Baseline 的对比测试。

报告分工

- 张翔负责 Basic idea、部分 Detailed algorithm 的书写以及 Slide 的制作。
- 李家昊负责 Detailed algorithm 部分的书写。
- 宋浩轩负责 Stability 测试部分以及 Contribution 部分的书写。
- 李浩源负责 Baseline 测试部分的书写。

参考文献

- [1] Yaroslav Ganin and Victor S. Lempitsky. Unsupervised domain adaptation by backpropagation. *ArXiv*, abs/1409.7495, 2015.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [3] Yue He, Zheyang Shen, and Peng Cui. Towards non-i.i.d. image classification: A dataset and baselines, 2019.
- [4] Claudio Michaelis, Ivan Ustyuzhaninov, Matthias Bethge, and Alexander S. Ecker. One-shot instance segmentation. *ArXiv*, abs/1811.11507, 2018.
- [5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc., 2017.
- [6] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6450–6458, 2017.