The Many Faces of Robustness: A Critical Analysis of OOD Generalization

Inspiration:

1. OOD robustness不应该是一个简单的指标，它在不同distribution shift下应该是不同的指标。难以推断现有方法中哪个能更广泛地work。 做实验时，最好在多个不同的OOD数据集上，验证方法在不同类分布偏移下的效果。本文建议：至少要在ImageNet-C和ImageNet-R上做实验。
2. 模型有使用texture做预测的趋势。使用数据增强破坏texture可能有助于OOD。

Four datasets

ImageNet-Renditions (ImageNet-R)

包含200个类。有点类似PACS，有多种风格，只不过更多。

StreetView StoreFronts (SVSF)

分布变换包括：country、year、camera。

DeepFashion Remixed

由于相机变化引起的分布偏移。包括：物体大小、物体遮挡关系、相机角度、相机变焦等。

DeepAugment

image-to-image地生成augmented images。使用一个网络，在网络的不同层加入随机挑选的增强函数，包括：zeroing, negating, convolving, transposing, applying activation functions 等等.

Experiments

ImageNet-Renditions (ImageNet-R)

类似PACS，风格变换为主

	ImageNet-200 (%)	ImageNet-R (%)	Gap
ResNet-50	7.9	63.9	56.0
+ ImageNet-21K *Pretraining* (10× labeled data)	7.0	62.8	55.8
+ CBAM (*Self-Attention*)	7.0	63.2	56.2
+ ℓ_∞ Adversarial Training	25.1	68.6	43.5
+ Speckle Noise	8.1	62.1	54.0
+ Style Transfer Augmentation	8.9	58.5	49.6
+ AugMix	7.1	58.9	51.8
+ DeepAugment	7.5	57.8	50.3
+ DeepAugment + AugMix	8.0	53.2	45.2
ResNet-152 (*Larger Models*)	6.8	58.7	51.9

Conclusions

1. pretrain几乎没用。（参考ImageNet-21的数据）
2. Self-attention会扩大IID-OOD gap。
3. AugMix和DeepAugment增强了IID和OOD，Style Transfer对IID有损害。→

StreetView StoreFronts

国家、相机、年份引起的变化。

		Hardware	Year		Location
Network	IID	Old	2017	2018	France
ResNet-50	27.2	28.6	27.7	28.3	56.7
+ Speckle Noise	28.5	29.5	29.2	29.5	57.4
+ Style Transfer	29.9	31.3	30.2	31.2	59.3
+ DeepAugment	30.5	31.2	30.2	31.3	59.1
+ AugMix	26.6	28.0	26.5	27.7	55.4

Conclusions

1. 数据增强对于这类变换不是很有效，因为类似建筑结构的变化是更高层次的的语义特征的变化，现有的数据增强不是很能捕捉这类变换。

DeepFashion Remixed

			Size		Occlusion		Viewpoint		Zoom	
Network	IID	OOD	Small	Large	Slight/None	Heavy	No Wear	Side/Back	Medium	Large
ResNet-50	77.6	55.1	39.4	73.0	51.5	41.2	50.5	63.2	48.7	73.3
+ ImageNet-21K *Pretraining*	80.8	58.3	40.0	73.6	55.2	43.0	63.0	67.3	50.5	73.9
+ SE (*Self-Attention*)	77.4	55.3	38.9	72.7	52.1	40.9	52.9	64.2	47.8	72.8
+ Random Erasure	78.9	56.4	39.9	75.0	52.5	42.6	53.4	66.0	48.8	73.4
+ Speckle Noise	78.9	55.8	38.4	74.0	52.6	40.8	55.7	63.8	47.8	73.6
+ Style Transfer	80.2	57.1	37.6	76.5	54.6	43.2	58.4	65.1	49.2	72.5
+ DeepAugment	79.7	56.3	38.3	74.5	52.6	42.8	54.6	65.5	49.5	72.7
+ AugMix	80.4	57.3	39.4	74.8	55.3	42.8	57.3	66.6	49.0	73.1
ResNet-152 (*Larger Models*)	80.0	57.1	40.0	75.6	52.3	42.0	57.7	65.6	48.9	74.4

Conclusions

1. 没有方法在OOD上有显著的提高。因此此数据集只有IID有参考价值。
2. 没有方法能在所有的OOD变换下始终保持高性能。

Overall Conclusions

1. 更大的模型和更多样的数据有助于降低texture bias。
2. 在ImageNet-C上work的方法，在许多real-world blurry图片上也work，说明ImageNet-C可以用于验证真实世界鲁棒性。