Adversarial and Isotropic Gradient Augmentation for Image Retrieval With Text Feedback

Fuxiang Huang, Lei Zhang, *Senior Member, IEEE*, Yuhang Zhou, and Xinbo Gao

用于文本反馈图像检索的对抗性和各向同性梯度增强算法

报告人：杨正颖、曾靖涵、万明扬

2024/5/15

重庆大学 CHONGQING UNIVERSITY

目录

重庆大学 CHONGQING UNIVERSITY

目录

引言| 研究背景

图像检索应用：产品搜索、人脸识别、人员重新识别、互联网搜索等

图像检索范式

- 图像 ⟶ 图像
- 文本 ⟶ 图像（跨模态）

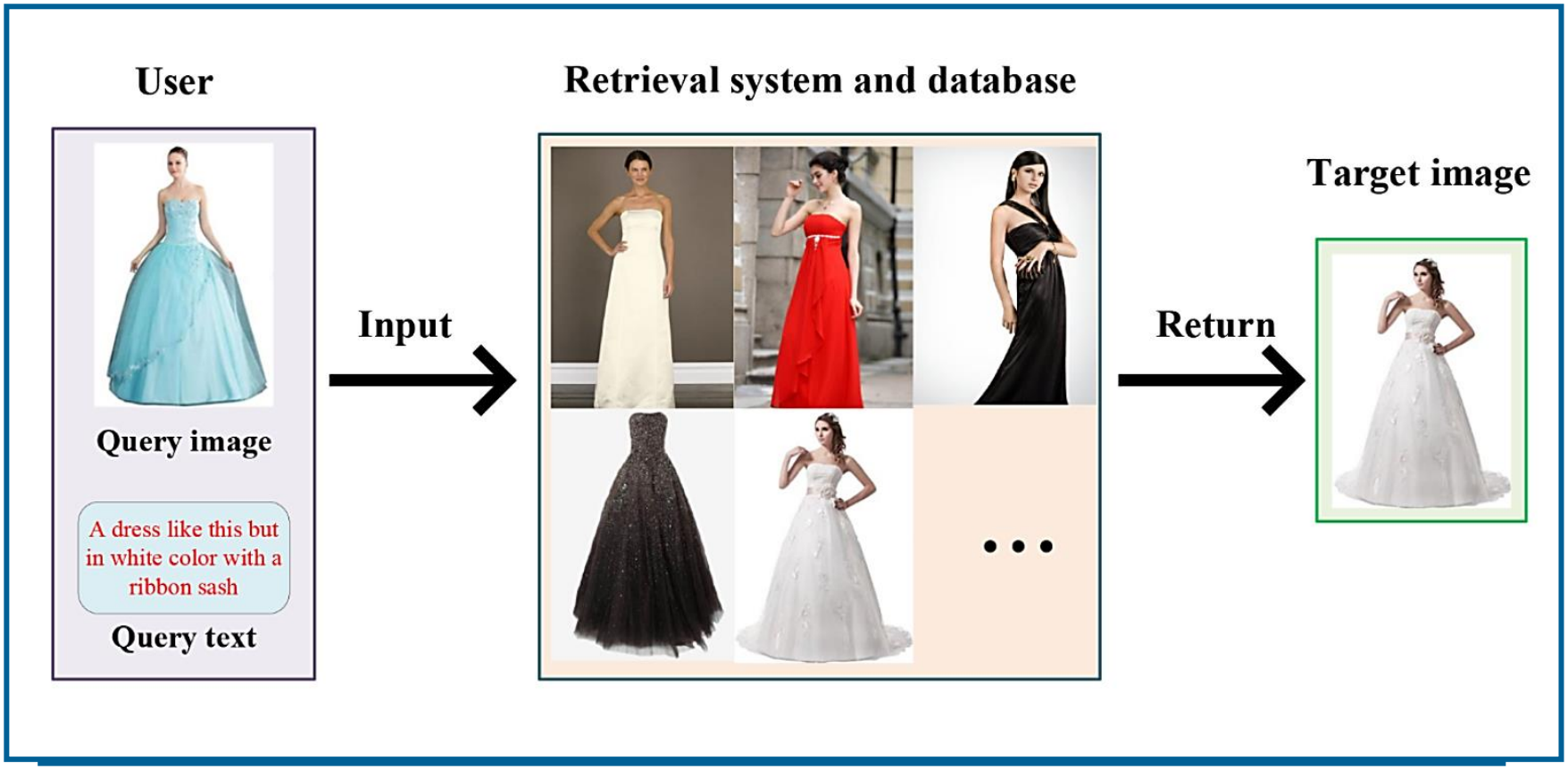
最大的挑战：准确理解用户意图

↓

带文本反馈的图像检索（IRTF）

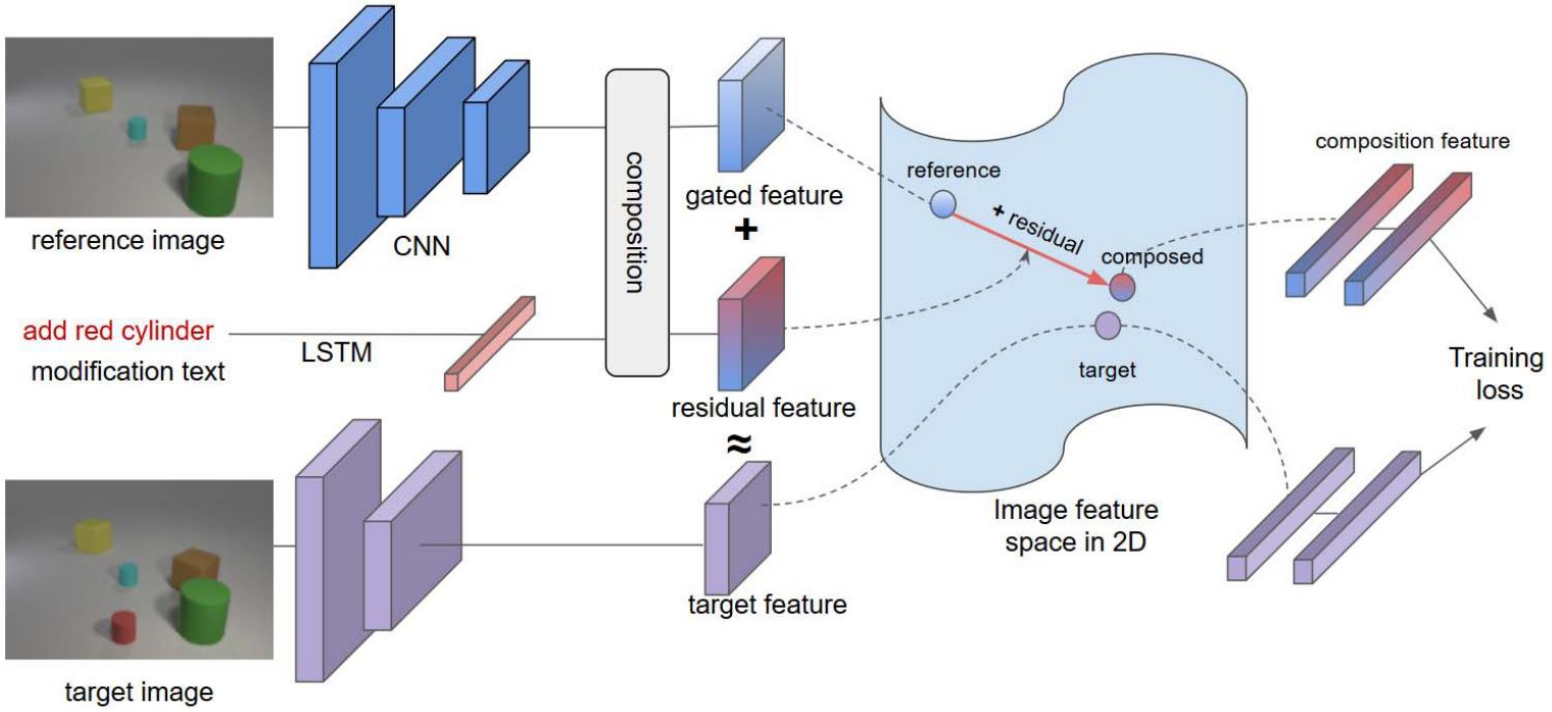
带文本反馈的图像检索（IRTF）：

范式：图像+文本 → 图像



带文本反馈的图像检索 （IRTF）方法:

- TIRG网络：获得查询图像和文本的特征组合 [22]；



[22]N. Vo, J. Lu, S. Chen, K. Murphy, and J. Hays, “Composing text and image for image retrieval - An empirical odyssey,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 6439–6448.

带文本反馈的图像检索 （IRTF）方法:

- 首先提出可以获得查询图像和文本特征组合的TIRG网络[22]；

改进方法之一

- ComposeAE模型[26]：查询图像和目标图像位于一个公共的复数空间中；

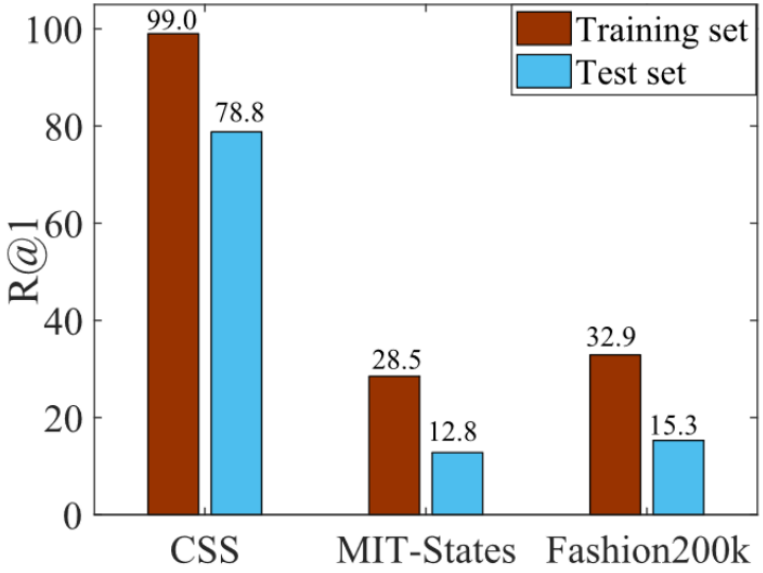
问题与挑战：容易过拟合、泛化性差。

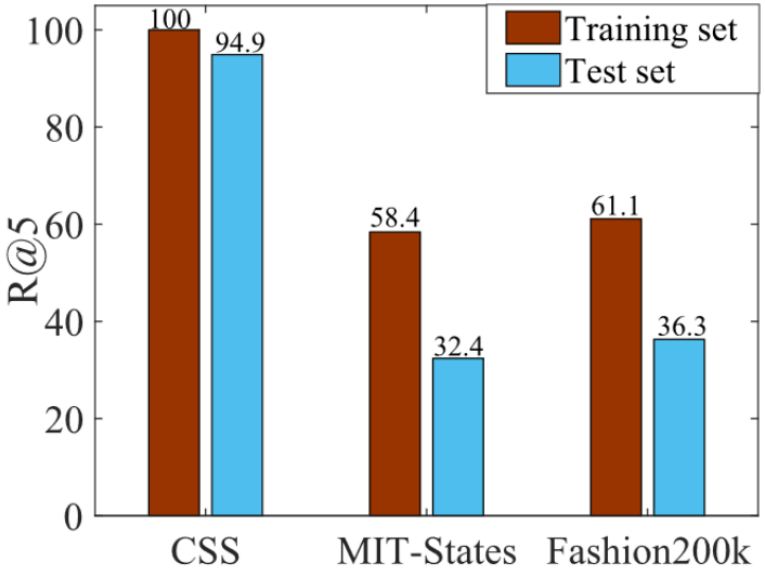
[22]N. Vo, J. Lu, S. Chen, K. Murphy, and J. Hays, “Composing text and image for image retrieval - An empirical odyssey,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 6439–6448.

[26]M. U. Anwaar, E. Labintcev, and M. Kleinsteuber, “Compositional learning of image-text query for image retrieval,” in Proc. Winter Conf. Appl. Comput. Vis., 2021, pp. 1140–1149.

原因：

- 过拟合：训练集样品数量不足，使训练模型时过拟合，导致测试性能不理想。
- 泛化性差：训练集分布多样性低；训练集和测试集的概率分布不同。





引言| 动机与想法

提高模型泛化能力的方法：

零样本学习：
在没有相应训练样本的情况下正确识别来自未见过测试类中的对象。

数据增强
从有限的训练数据中生成多样化和丰富的数据来**缓解过度拟合**并**促进模型泛化**。

对抗性数据增强：生成**对抗性样本**进行**对抗性训练**来增强数据。

域适应：
将知识从标记丰富的源域转移到未标记的目标域，以提高未标记目标域的泛化性。

需要训练额外的生成模型，且计算成本高。

重庆大学 CHONGQING UNIVERSITY

问题一：容易过拟合

目标：解决模型过拟合问题、提高IRTF模型鲁棒性

- **思路1：生成对抗性样本进行对抗性训练来增强数据**

  缺点：耗时、计算效率低

- **思路2：通过基于梯度的正则化可以达到数据增强的等效效果**

解决方法：提出了基于对抗梯度增强的正则化

问题二：泛化性差

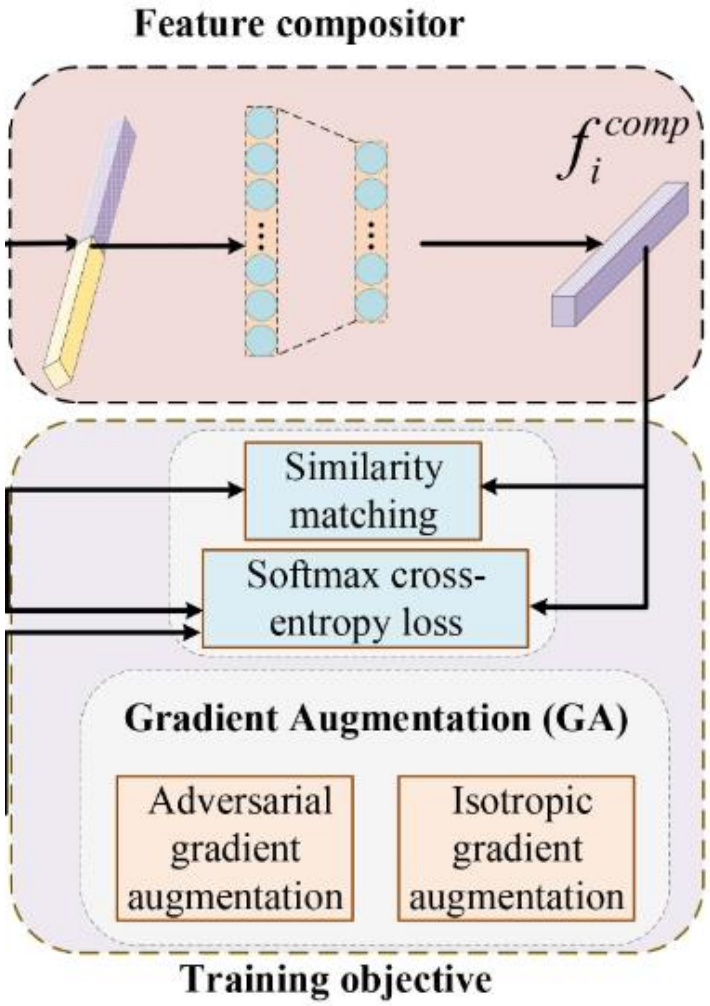
目标：解决训练集分布多样性不足的问题

- **思路1：将更多的样本输入模型，最终将反映在梯度的变化上**
- **思路2：具有不同分布的样本贡献了不同的梯度**

  梯度的合理调整可以提高训练分布的多样性
- **思路3：通过改变无穷范数球中的梯度来丰富训练分布的多样性**

解决方法：提出了基于各向同性梯度增强的正则化

引言| 贡献



梯度增强（GA）

- **显式对抗性梯度增强**
- **隐式各向同性梯度增强**

引入深度度量学习来训练模型

- **softmax交叉熵损失：**学习判别特征表示
- **相似性匹配损失：**
  使查询图像-文本组合表示与目标图像对齐

加权谐波均值(WHM)：评估模型的泛化。

引言| 贡献与新颖性

贡献1：提出了梯度增强（GA）的正则化方法

- **目的**：在不增加额外计算的情况下改进IRTF任务的模型泛化。

贡献2：从对抗训练的角度提出了显示对抗性梯度增强

- **目的**：缓解了模型过拟合的问题。 • **优点**：不需要任何对抗性样本

贡献3：推导出了一种隐式各向同性梯度增强

- **目的**：在不生成各种样本的情况下丰富训练集分布的多样性。

贡献4：提出了一种加权谐波均值(WHM)的合理评估协议

- **优点**：能够反映训练集和测试集的整体性能。

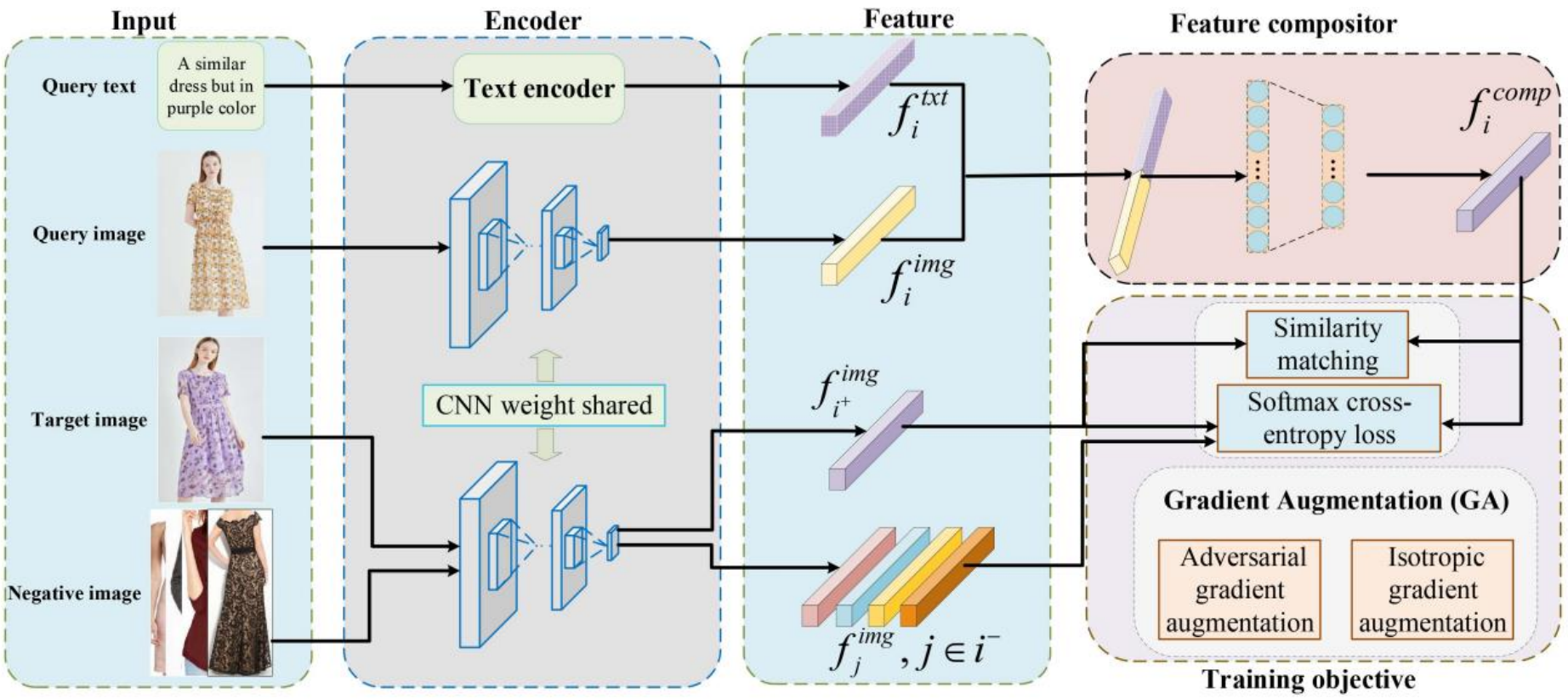
目录

引言

算法

实验

算法|网络框架



IRTF的任务：

$$\max_{\theta} \quad K(f_i^{comp}, f_{i^+}^{img})$$

学习相关特征

算法|显式对抗性梯度增强

（1）假定在样本x，加入δ微妙且难以察觉的对抗性扰动使得：

$$\|\delta\| \leq \in \text{和} c(x) \neq c(x+\delta)$$

（2）进行对抗训练的损失函数，可以根据FGSM [53]表示为：

$$\tilde{L}(x) = \frac{1}{2}(L(x) + L(x+\delta))$$

（3）经过泰勒展开式：

$$\tilde{L}(x) = L(x) + \frac{\delta}{2} * \frac{\partial L(x)}{\partial x} + o(\|\delta\|)_p$$

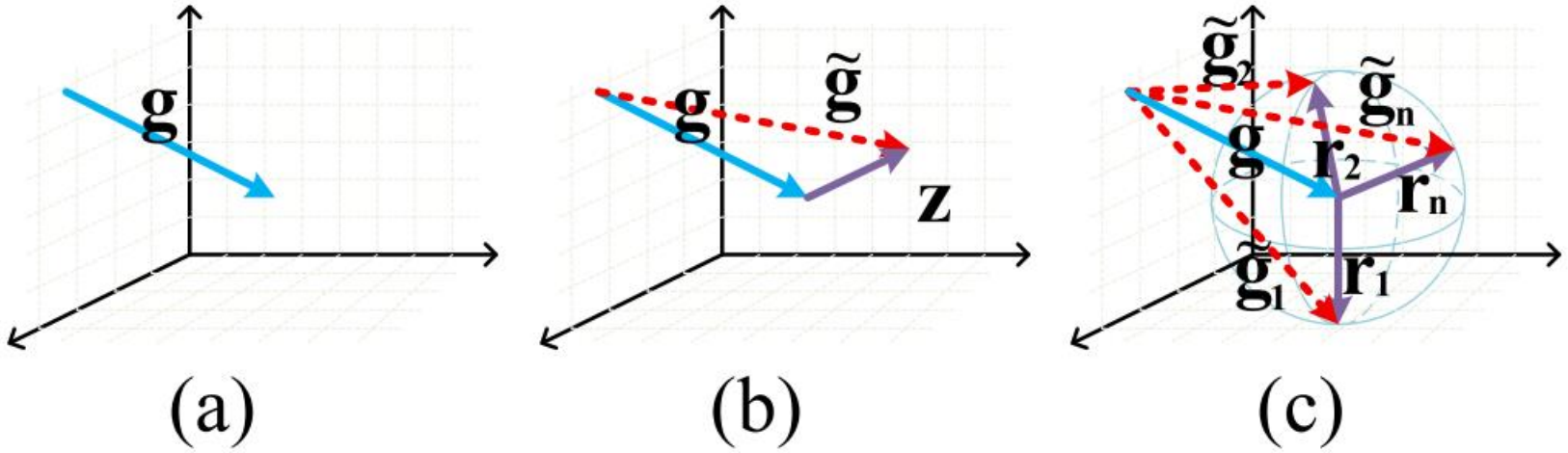
（4）令q 为p 的对偶范数：

对抗性梯度增强目标

$$\frac{\partial L(x)}{\partial x} = \in \left\| \frac{\partial (x)}{\partial x} \right\|_q = L_{GA}^A$$

[53] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in Proc. Int. Conf. Mach. Learn. Representations, 2015

算法|隐式各向同性梯度增强



$$L_{GA}^{I} = \gamma \int r d\theta = \gamma \ r.\theta$$

通过在梯度上添加一个各向同性的随机向量，使梯度在球体范围内随机扩散，从而获得更广泛的优化方向。

算法|训练目标

在训练阶段，对于第 i 个查询，我们创建一个由 1 个正例$f_{i^+}^{img}$组成的集合 N_i和K个负面例子，采用 softmax 交叉熵损失作为标准损失L_{CE}

$$L_{CE} = \frac{1}{B}\sum_{i=1}^{B} \log\{1 + \exp(k\left(f_i^{comp}, f_{i^-}^{img}\right) - \left(f_i^{comp}, f_{i^+}^{img}\right))\}$$

为了使查询复合表示与目标图像表示对齐,引入了相似度匹配损失L_S

$$L_S = \frac{1}{B}\sum_{i=1}^{B} \left\| f_i^{comp} - f_{i^+}^{img} \right\|^2$$

为了减轻过度拟合并提高训练分布的多样性，将梯度增强添加到训练目标中。

算法|未知测试

总损失计算公式： $L_{all} = L_{CE} + \eta L_S + L_{GA}^I + L_{GA}^A$

根据域适应（DA）[58]理论，学习模型从源域到目标域的泛化界限显示预期测试误差的R^t (h)上限如下

$$R^t(\mathrm{h}) \le R^s(\mathrm{h}) + d_H(D_s, D_t) + \lambda$$

目标域应该是已知的(但我们测试集是未知的)

$$R^t(\mathrm{h}) \le \sum_{i=1}^{N_s} \pi_i R_i^s(\mathrm{h}) + \frac{\rho}{2} + \lambda_\pi$$

基于领域适应理论的领域泛化（DG） [59]修正理论形式化为未见过的测试数据

[58] S. Ben-David et al., "A theory of learning from different domains," Mach. Learn., vol. 79, no. 1, 2010, Art. no. 151C175.

[59] I. Albuquerque, J. Monteiro, and M. Darvishi, "Adversarial targetinvariant representation learning for domain generalization," 2019,arXiv:1911.00804v3.

模型和泛化界限理论之间的联系

总损失计算公式： $L_{all} = L_{CE} + \eta L_S + L_{GA}^I + L_{GA}^A$

- **联系1：源误差，L_{CE}, L_S 可以促进最小化源误差，L_{GA}^I, L_{GA}^A 本质上解决了两个来源的泛化问题，即原始训练数据与对抗性样本，可以进一步有助于最小化中源误差的凸组合**
- **联系2：最大分布差异，通过各向同性梯度增强的幅度可以通过γ来控制，源之间的分布差异也可以是有界的**
- **联系3：多源联合误差，理想假设的联合误差λ_π，通常被认为足够小，因为存在源的标签监督**

目录

实验|数据集

数据集设置

TABLE I
DATASET STATISTICS

	CSS	MIT-States	Fashion200k
Train queries	19012	43207	172049
Test queries	19057	82732	33480
Average of target images per query	1	26.7	3
Batch size	200	32	32

CSS数据集：包含大小、颜色信息的三位立方体　　拆分遵循TIRG的规范

MIT数据集：形容词（状态）+名词（事物）

Fashion200k：衣服

实验|对比方法

方法	图像特征提取模块	文本特征提取模块
Show and Tell Parameter Hashing Attribute as Operator Relationship FiLM CoSMo TIRG TIRG-DIM	ResNet-18	LSTM
CQIRULBF	Faster-RCNN	TEP
JVSM VAL	MobileNet	LSTM
ComposeAE	ResNet-18	BERT

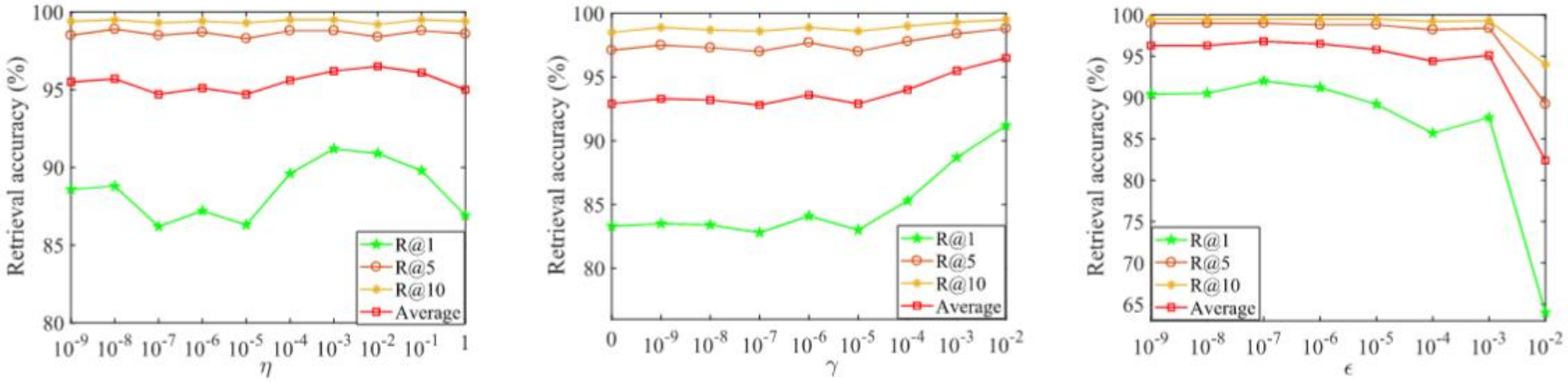
实验|超参数设置

损失函数的超参数设置

$$\mathcal{L}_{all} = \mathcal{L}_{CE} + \eta\mathcal{L}_S + \mathcal{L}_{GA}^{A} + \mathcal{L}_{GA}^{I}$$

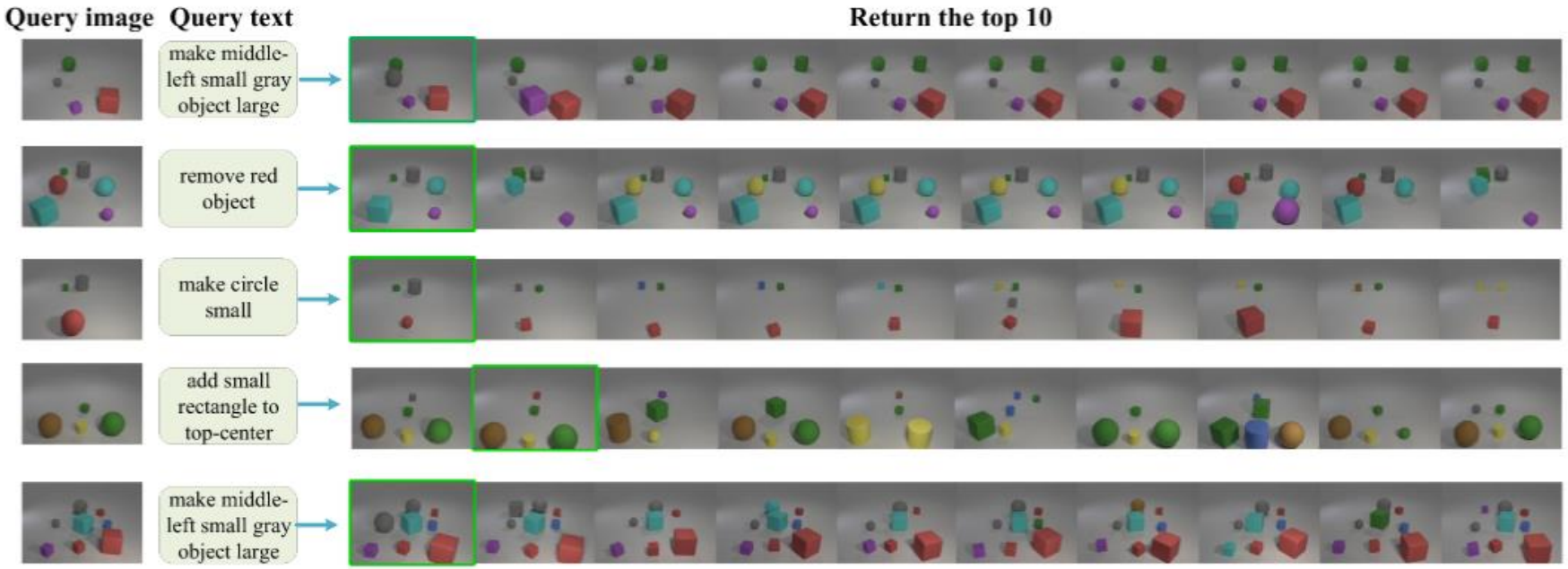
$$\mathcal{L}_{GA}^{A} = \epsilon \left\| \frac{\partial \mathcal{L}}{\partial x} \right\|_q \qquad \mathcal{L}_{GA}^{I} = \gamma \int \mathbf{r} d\boldsymbol{\Theta} = \gamma \mathbf{r} \cdot \boldsymbol{\Theta}$$

10^{-6} 10^{-3} 10^{-2}



超参数的敏感性

实验|效果

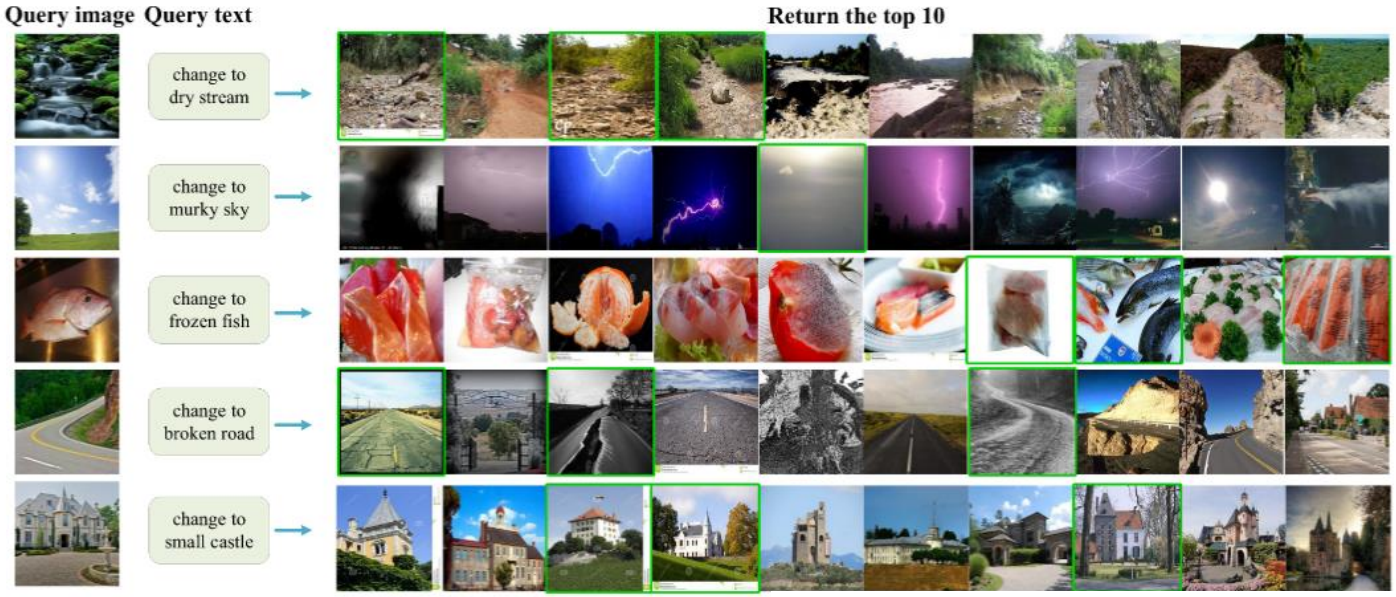


CSS数据集上的定性效果

CSS数据集上的定量效果

Method	R@1	R@5	R@10	Average
Concatenation	60.6	88.2	92.8	80.5
Show and Tell	33.0	75.0	83.0	63.7
Para.Hasing	60.5	88.1	92.9	80.5
Relationship	62.1	89.1	93.5	81.6
Film	65.6	89.7	94.6	83.3
TIRG	78.8	94.9	97.3	90.3
TIRG-DIM	77.0	95.6	97.6	90.1
CQIRULBF	79.2	–	–	–
TIRG+GA	**91.2**	**98.8**	**99.5**	**96.5**

实验|效果



MIT-States数据集上的定性效果

MIT-States数据集上的定量效果

Method	R@1	R@5	R@10	Average
Concatenation	11.8	30.8	42.1	28.2
Show and Tell	11.9	31.0	42.0	28.3
Att. as Operator	8.8	27.3	39.1	25.1
Relationship	12.3	31.9	42.9	29.0
Film	10.1	27.7	38.3	25.4
TIRG	12.2	31.9	42.9	29.0
TIRG-BERT	13.3	34.5	46.8	31.5
TIRG-DIM	14.1	33.8	45.0	31.0
CQIRULBF	14.7	35.3	46.6	32.2
ComposeAE	13.9	35.3	47.9	32.4
TIRG+GA	13.6	32.4	43.2	29.7
TIRG-BERT+GA	**15.4**	36.3	47.7	**33.2**
ComposeAE+GA	14.6	**37.0**	**47.9**	**33.2**

实验|效果

Fashion200k数据集上的定性效果

Fashion200k数据集上的定量效果

提升不明显 ⟶

Method	R@1	R@10	R@50	Average
Concatenation	11.9	39.7	62.6	38.1
Show and Tell	12.3	40.2	61.8	38.1
Param Hashing	12.2	40.0	61.7	38.0
Relationship	13.0	40.5	62.4	38.6
Film	12.9	39.5	61.9	38.1
TIRG	15.3	44.3	65.0	41.5
TIRG-BERT	19.0	49.1	74.4	47.5
TIRG-DIM	17.4	43.4	64.5	41.8
JVSM	19.0	52.1	70.0	47.0
VAL	22.9	50.8	72.7	48.8
CQIRULBF	17.8	48.4	68.5	44.9
ComposeAE	22.4	55.0	71.6	49.7
CoSMo	23.3	50.4	69.3	47.7
ComposeAE+GA	25.2	52.8	71.2	49.7
TIRG+GA	24.0	**57.2**	75.7	52.3
TIRG-BERT+GA	**31.4**	54.1	**77.6**	**54.4**

实验|WHM

加权调和平均数 $WHM = \frac{(1+\alpha) * M_s * M_t}{\alpha * M_t + M_s}$

Ms-训练集上的精度

Mt-测试集上的精度

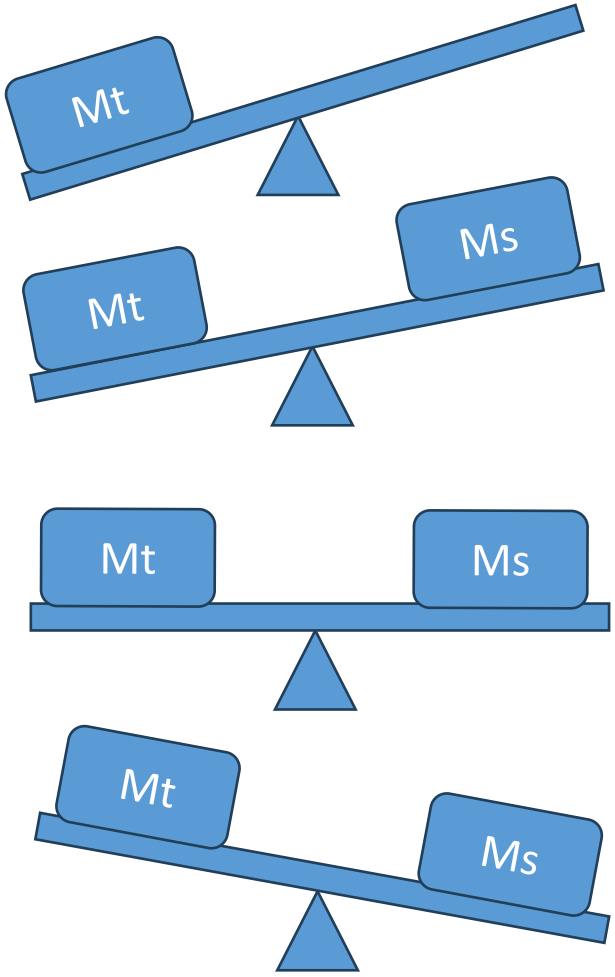
α-标量，训练集的重要性

α=0， WHM = Mt

0<α<=1

α=1， WHM = $\frac{Ms+Mt}{2}$

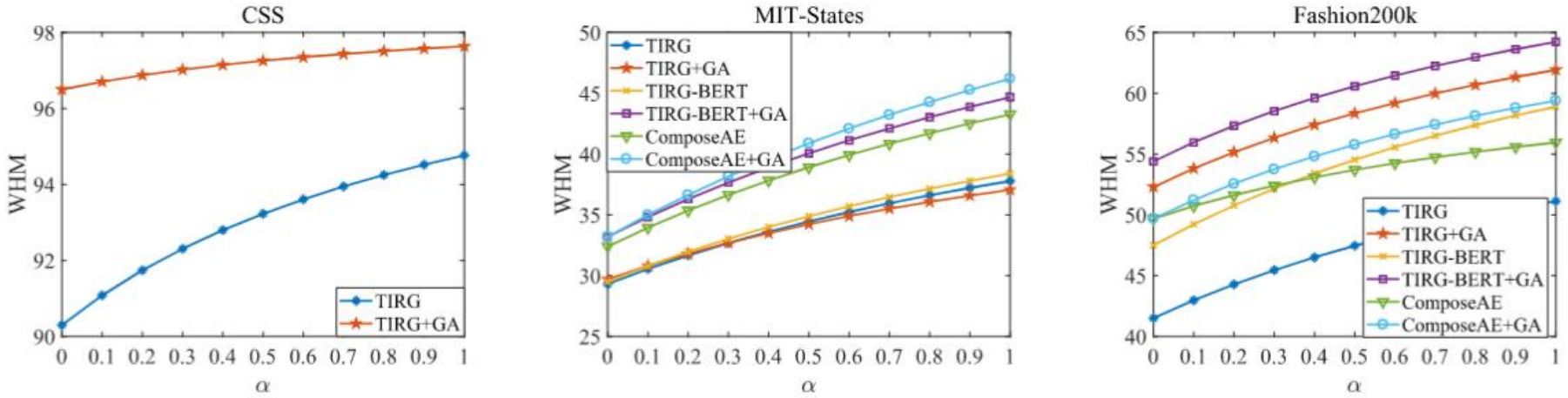
α>1



实验|WHM

Dataset	CSS			MIT-States			Fashion200k		
Method \ Metric	M_s	M_t	WHM ($\alpha = 0.5$)	M_s	M_t	WHM ($\alpha = 0.5$)	M_s	M_t	WHM ($\alpha = 0.5$)
TIRG	**99.7**	90.3	93.2	**53.2**	29.3	**34.5**	66.6	41.5	47.5
TIRG+GA	98.8	**96.5**	**97.3**	49.3	**29.7**	34.2	**75.9**	**52.3**	**58.3**
TIRG-BERT	–	–	–	55.0	29.5	34.9	77.5	47.5	54.5
TIRG-BERT+GA	–	–	–	**68.3**	**33.2**	**40.1**	**78.4**	**54.4**	**60.6**
ComposeAE	–	–	–	65.1	32.4	38.9	64.0	49.7	53.7
ComposeAE+GA	–	–	–	**76.0**	**33.2**	**40.9**	**73.8**	**49.7**	**55.8**

WHM ↑ 指标对比



WHM分数中α的敏感性

实验|梯度增强分析

Method	Training time	R@1	R@5	R@10	Average
Baseline	111.28 s/epoch	78.8	94.9	97.3	90.3
Baseline+FGSM	248.57 s/epoch	79.7	96.4	98.1	91.4
Baseline+$\mathcal{L}_{GA}^{A}$	192.54 s/epoch	**82.8**	**97.4**	**98.6**	**92.9**

对抗性梯度增强和传统对抗性学习的比较　　Baseline：TIRG

Method	R@1	R@5	R@10	Average
Baseline	77.9	94.3	97	89.7
Baseline+$\mathcal{L}_1$	81.4	96.3	98.2	92.0
Baseline+$\mathcal{L}_2$	78.8	94.9	97.3	90.3
Baseline+$\mathcal{L}_{GA}^{I}$	**88.8**	**98.8**	**99.5**	**95.7**

各向同性梯度增强和正则化的比较

$$\mathcal{L}_{GA}^{I} = \gamma \int \mathbf{r} d\mathbf{\Theta} = \gamma \mathbf{r} \cdot \mathbf{\Theta}$$

$$L_{\mathrm{L1}} = L_{\mathrm{data}} + \lambda \sum_{i=1}^{n} |w_i| \qquad L_{\mathrm{L2}} = L_{\mathrm{data}} + \lambda ||\mathbf{w}||_2^2$$

实验|消融实验

Loss function	R@1	R@5	R@10	Average
$\mathcal{L}_{CE}$	78.8	94.9	97.3	90.3
$\mathcal{L}_{CE} + \mathcal{L}_{S}$	81.0	96.4	98.2	91.9
$\mathcal{L}_{CE} + \mathcal{L}_{GA}^{A}$	82.8	97.4	98.6	92.9
$\mathcal{L}_{CE} + \mathcal{L}_{GA}^{I}$	87.8	98.1	99.1	95.0
$\mathcal{L}_{CE} + \mathcal{L}_{S} + \mathcal{L}_{GA}^{A}$	83.0	97.1	98.5	92.9
$\mathcal{L}_{CE} + \mathcal{L}_{S} + \mathcal{L}_{GA}^{I}$	90.2	99.1	99.6	96.3
$\mathcal{L}_{CE} + \mathcal{L}_{GA}^{A} + \mathcal{L}_{GA}^{I}$	89.1	98.8	99.5	95.8
$\mathcal{L}_{CE} + \mathcal{L}_{S} + \mathcal{L}_{GA}^{A} + \mathcal{L}_{GA}^{I}$	**91.2**	**98.8**	**99.5**	**96.5**

消融实验

THANKS

汇报人：杨正颖、曾靖涵、万明扬

2024/5/15