Compound Projection Learning for Bridging Seen and Unseen Objects

Wenli Song, Lei Zhang* and Xinbo Gao

微电子与通信工程学院

阅读分享汇报人：丁睿、冯浩

第四组

目录

CONTENTS

引言

一、需要解决的问题

目标识别需要大量标记数据来识别每个类别，大多数分类模型依赖于大量标记图像来获取有效的深度神经网络。随着图像数据的快速增长，有大量的对象缺乏足够的图像数据，收集每个类别的大量标记图像成本高昂且不切实际。特别是对于那些罕见和细粒度的对象，获取足够数量的标记图像非常困难和昂贵。为了解决这个问题，**迁移学习、少样本学习（FSL）和零样本学习（ZSL）**逐渐成为机器学习和计算机视觉的研究重点之一。此外，基于ZSL还研究了许多新任务，如图像检索和长尾学习。

二、相关背景

零样本学习（ZSL）考虑了一种测试数据在训练阶段完全不可用的极端情况即训练（已见）类和测试（未见）类是不相交的。一般的ZSL方法旨在学习视觉和语义空间之间的投影函数，基本方法是将语义空间作为嵌入空间，将视觉特征投影到其中，一旦建立了连接，未见样本就可以通过在嵌入空间中进行最近邻搜索来进行分类，由于视觉特征的维度远高于语义特征的维度，因此在语义空间中最近邻搜索到的可能是一些与视觉空间中不相关的元素，这个问题被称为枢纽点问题。为了克服这个问题，他们提出将类原型嵌入到视觉空间中，除此之外，还有一些方法提出利用潜空间作为嵌入空间，通常情况下这些嵌入方法在投影函数和嵌入空间上有所不同。

引言

零样本学习中的两个关键问题

① **投影偏差问题。**由于已见和未见类是不相交的类别，它们可能具有不同的视觉分布，因此，由已见类学习到的投影函数在应用于未见样本时可能存在语义偏差。如图1(a)所示，老虎和斑马分别是已见和未见类，我们观察到尽管老虎和斑马具有相似的属性表示，但它们的视觉分布显示出明显的差异，因此，当利用投影函数对未见类进行分类时，必然会引起投影偏差问题，零样本分类结果的有效性显然会降低。

② **缺乏对类别间相似关系的探索。**嵌入方法旨在对齐视觉和语义模态，但忽略了类别间的关系。事实上，语义嵌入表征了更具辨别性的信息，并且能够表示更复杂的数据结构，即语义流形，语义流形揭示了数据更准确的内在结构，使得语义表示与视觉特征和标签空间密切相关。

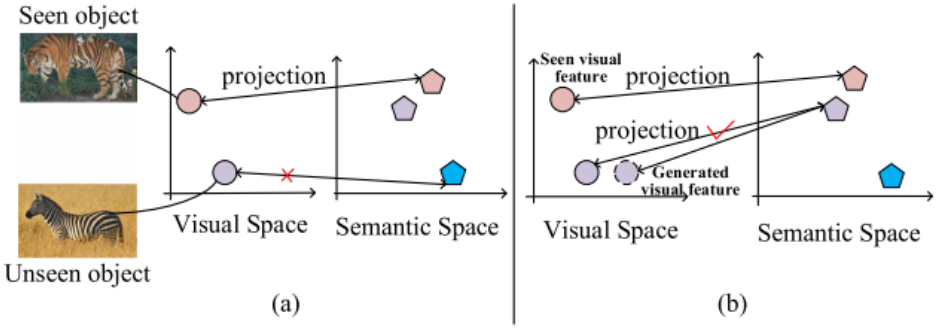


图1. 投影偏差问题示例（a）和我们的提议（b）。老虎和斑马分别是已见和未见类，当将在已见类上学到的投影函数应用于未见类时，由于视觉上的差异性，视觉特征被投影到语义空间中的不同位置，这就是偏差问题。

引言

为了解决由于不同视觉分布引起的投影偏差问题，进一步探索类别间的关系，受语义流形的启发，类似的视觉特征意味着类似的语义，训练相似性网络，该论文主要贡献可以总结如下：

1）提出了一个语义-视觉复合投影网络与生成模型合作，旨在构建已见和未见类之间的桥梁，其中利用了标记的已见样本和生成的未见样本来构建语义-视觉空间的连接。

2）受语义流形的启发，利用语义相似性来监督类间关系的学习，在类间相似性的监督下，知识可以在训练阶段从已见类转移到相似的未见类。

3）在基准数据集上进行了实验研究，并与最先进的模型进行了比较，以展示CPL模型的优越性。

相关工作

1） **语义嵌入空间**：语义嵌入空间是一个高维向量空间，使得已见和未见类别能够共享信息。由于依赖人工对特性进行注释，许多模型使用辅助文本语料库来构建语义空间，具有相似含义的单词可以在单词向量空间中投影到相似的位置。

2） **相似性网络**：相似性网络广泛应用于少样本学习。RN提出将支持集和查询集的视觉特征连接起来，然后计算关系分数，SalNet 基于RN提出了一种基于数据幻觉的网络。

3） **深度生成模型**：大多数生成式零样本学习方法基于生成对抗网络（GAN）和变分自编码器（VAE）。在文献[36]中，f-CLSWGAN被提出利用未见类别的语义特征和高斯噪声基于WGAN生成未见样本。

4） **半监督零样本学习模型**：为了解决对已见类别的模型偏差问题，一些模型被提出来利用未标记的未见样本来增强数据集，TMV 是一个传统的多视角嵌入框架。在文献[15]中，提出了一个基于无监督域自适应的正则化稀疏编码框架。

相关工作

5） **视觉-语义嵌入：**近年来提出了各种零样本学习方法，视觉空间用于描述样本的视觉外观内容，如何学习和构建视觉-语义嵌入在零样本学习方法中起着重要作用，早期的零样本学习方法大多集中于学习在视觉和语义空间之间建立映射，以将已见类别的知识转移到未见类别。例如，在文献[5]中，首次提出了DAP（直接属性预测）模型和IAP（间接属性预测）模型在属性分类器上学习，考虑到由于识别性能差而导致属性分类器的不可靠性，提出了许多零样本学习方法来学习标签嵌入。

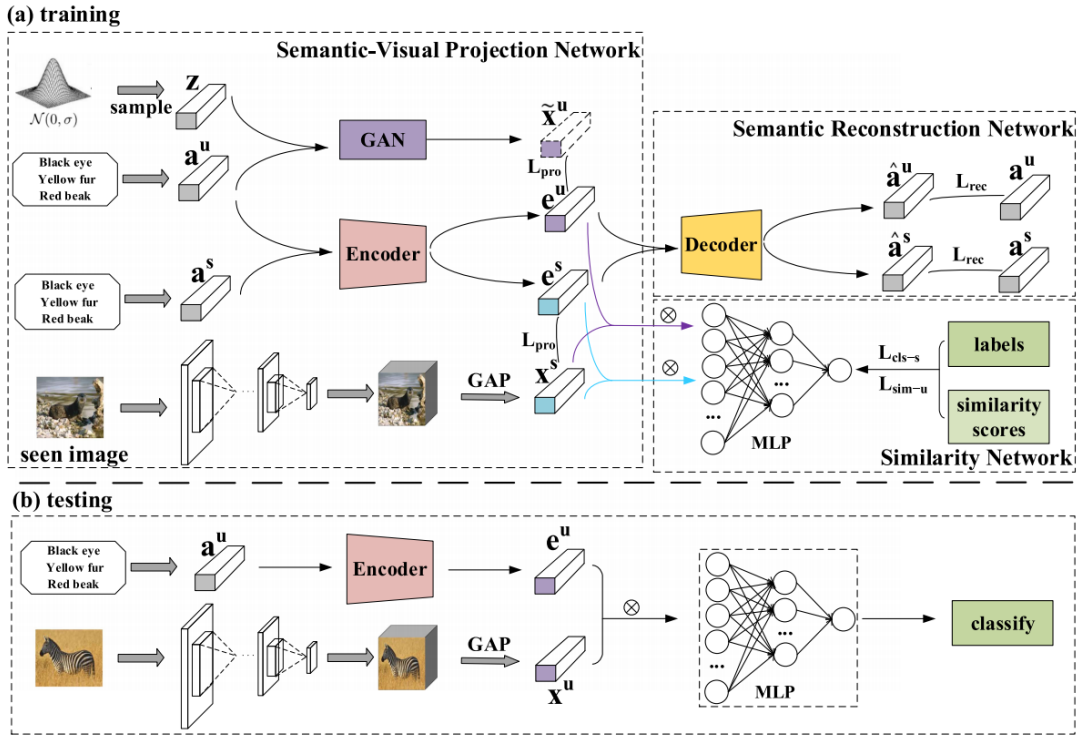


图 2. 训练框架 (a) 和测试框架 (b)。在训练阶段，我们首先利用生成对抗网络合成类别无关的样本。然后利用已见样本和这些生成的类别无关样本来学习语义-视觉投影网络，以减轻投影偏差问题。

目录

CONTENTS

研究方法 | 1. 问题定义

Notations	Meaning
c^s	可见类别的数量
N_s	有标签的可见类别的样本数量
c^u	不可见类别的数量
N_u	无标签的不可见类别的数量
$\mathbf{X}^s \in R^{N_s \times d}$	可见类别样本的**视觉特征**
$\mathbf{X}^u \in R^{N_u \times d}$	不可见类别样本的**视觉特征**
d	视觉特征的维度
$\mathbf{A}^s \in R^{N_s \times m}$	可见类别样本的**语义特征**
$\mathbf{A}^u \in R^{N_u \times m}$	不可见类别样本的**语义特征**
m	语义特征的维度
$\mathbf{Z}^s, \mathbf{Z}^u$	可见和不可见类别的**原型语义表示**
$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_s}] \in R^{N_s \times c^s}$	可见类别样本的标签
$\mathbf{y}_i$	对于第 i 个可见样本的独热编码的标签

零样本识别的设置：

$$\mathbf{Z}^s \cap \mathbf{Z}^u = \emptyset$$

可见和未见类别不相交

零样本学习ZSL的目标：

$$\mathcal{X} \rightarrow \mathcal{Y}^u$$

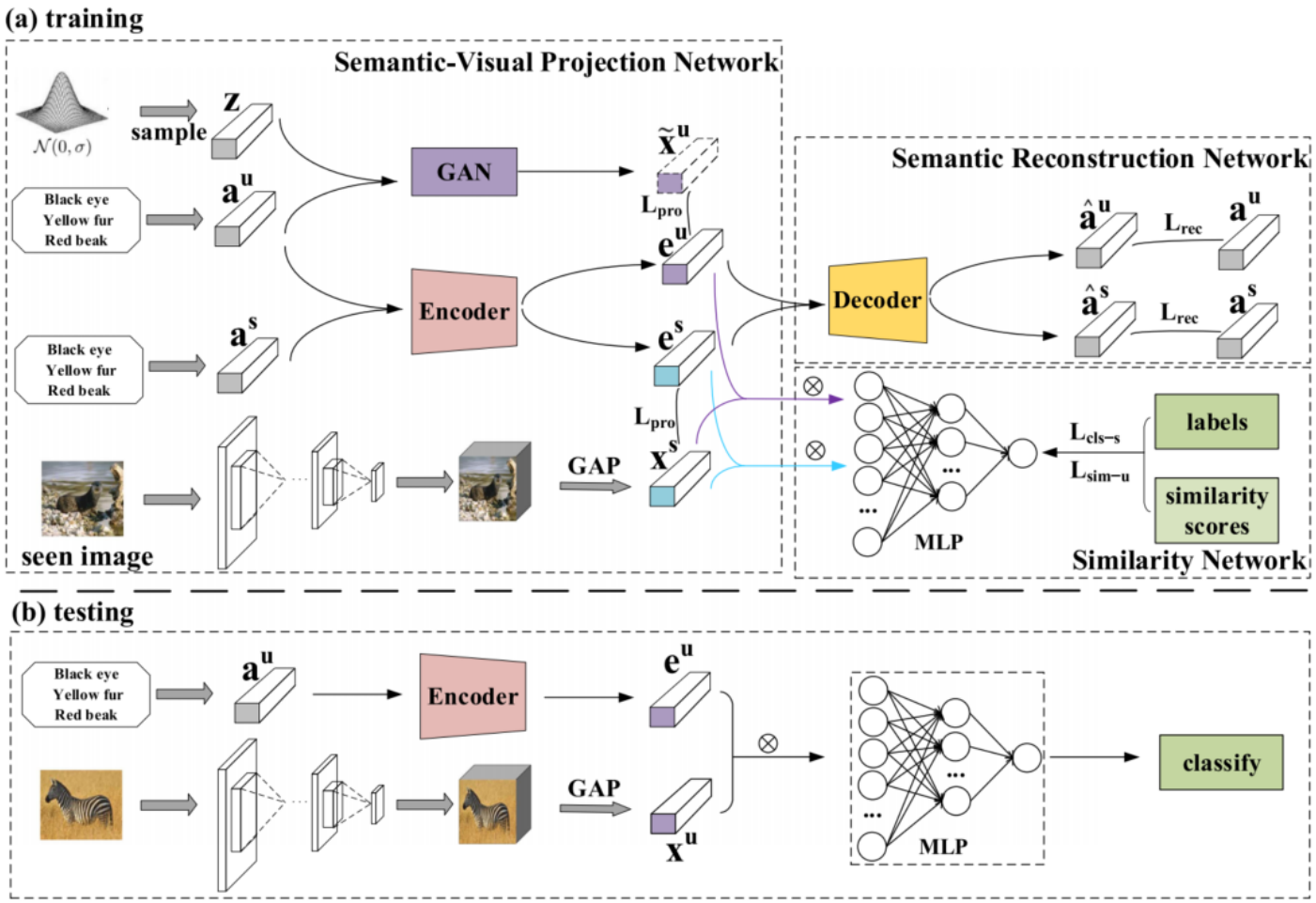
预测不可见类别样本的标签

广义零样本学习GZSL的目标：

$$\mathcal{X} \rightarrow \mathcal{Y}^s \cup \mathcal{Y}^u$$

预测所有类别样本的标签

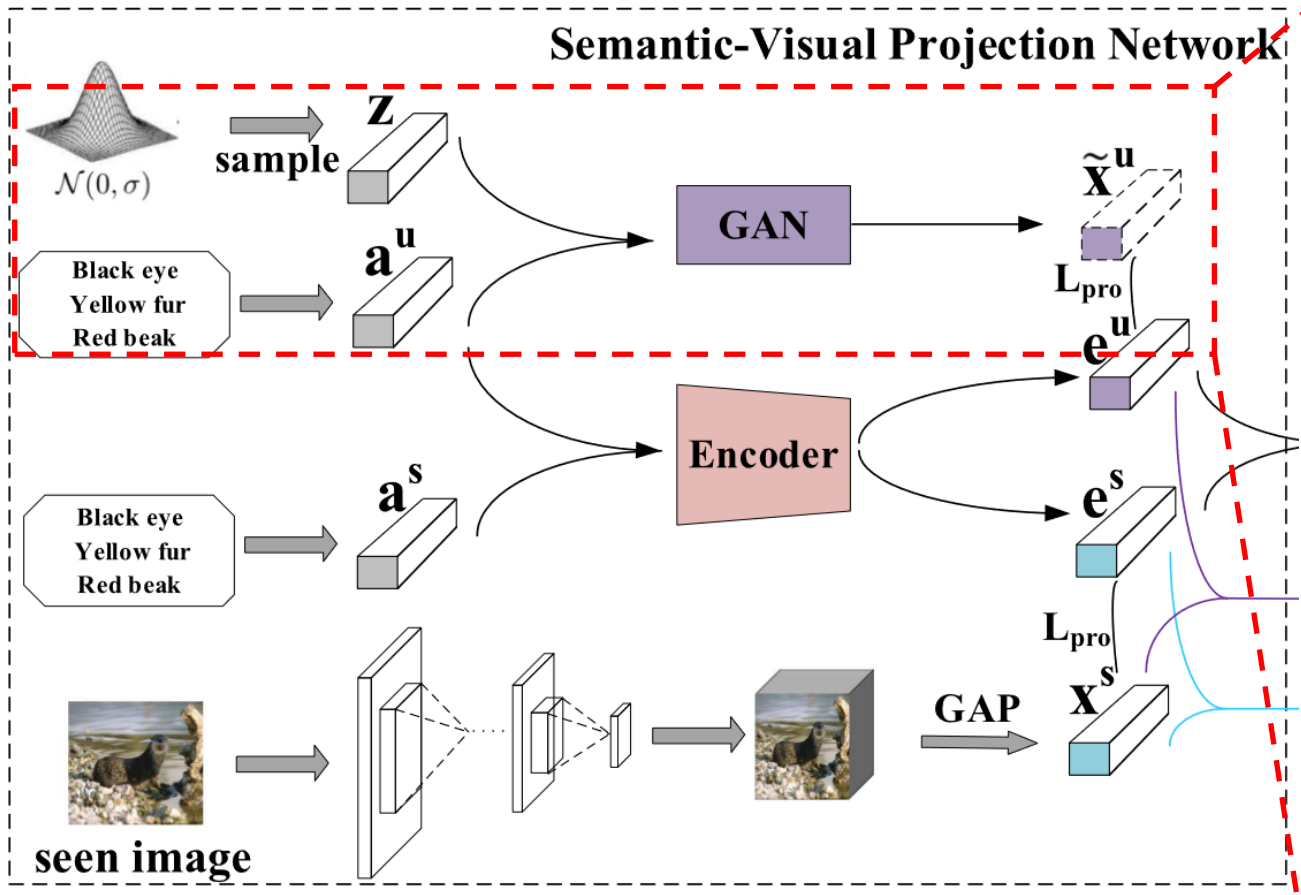
研究方法 | 1. 问题定义



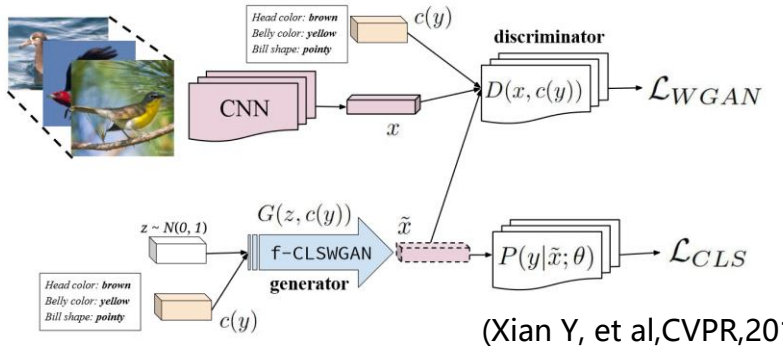
问题一：
投影偏移问题导致分类效果下降

问题二：
类间关系忽略导致语义混淆

研究方法 | 2. 语义-视觉复合投影网络



2.1 不可见类别的视觉特征生成



(Xian Y, et al,CVPR,2018)

生成器G

输入：**语义特征**和采样的**随机表示**
输出：**生成对应类的视觉特征**

判别器D

输入：**生成的视觉特征**和**真实的视觉特征**
输出：一个**实数值**用于计算损失

损失函数

$$\min_G \max_D \mathcal{L}_{WGAN} + \beta \mathcal{L}_{CLS}$$

×使用生成样本直接训练分类器

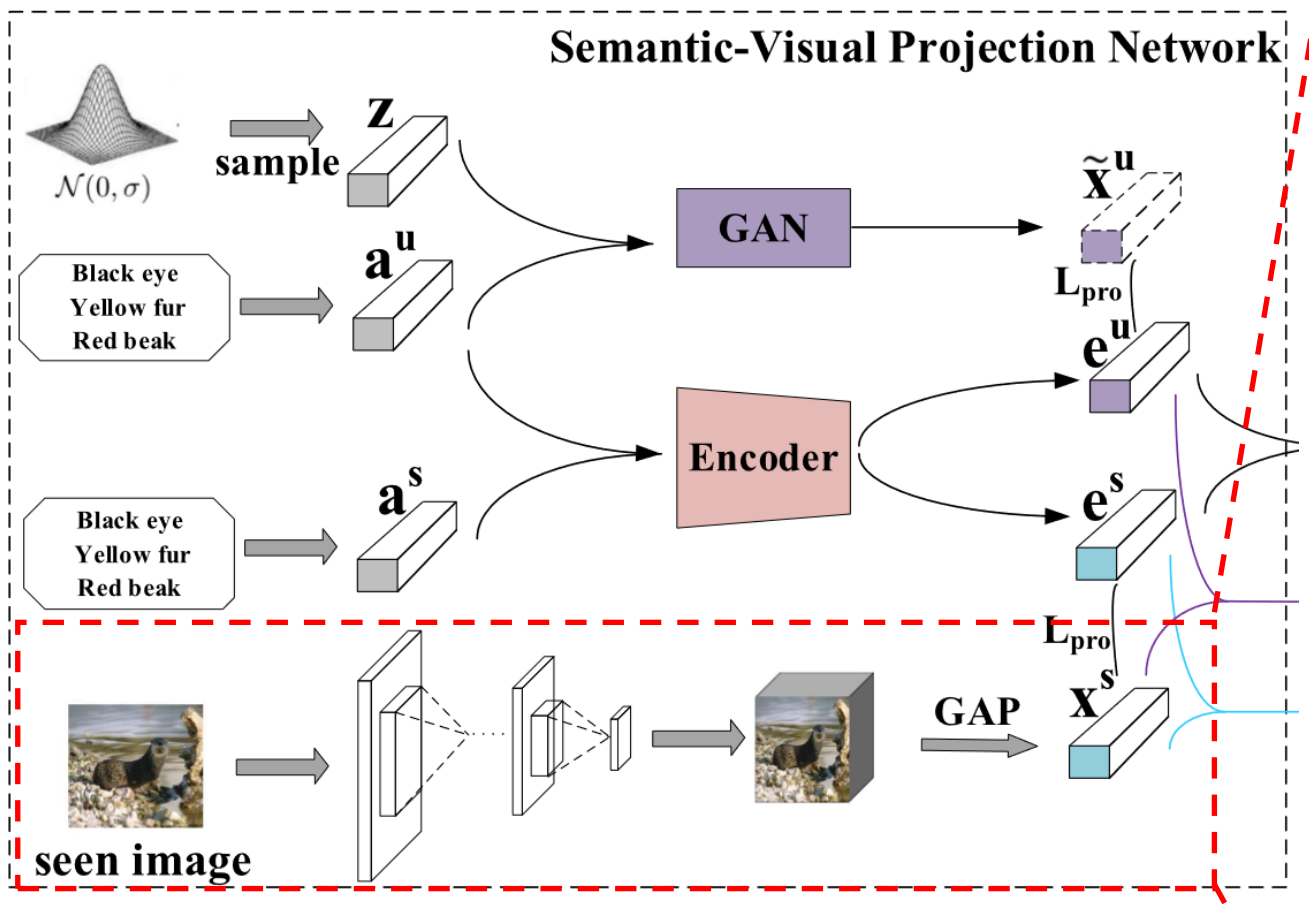
样本低质量、训练不稳定

√训练语义-视觉投影编码器

更有效、更可行
语义和视觉特征的映射

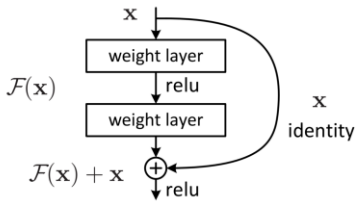
Xian Y, et al. Feature generating networks for zero-shot learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018

研究方法 | 2. 语义-视觉复合投影网络



2.2 可见类别的视觉特征表示

ResNet101 [1]

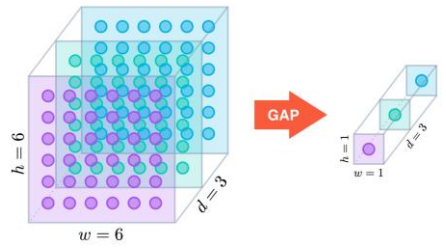


$$y = \mathcal{F}(x, \{W_i\}) + x.$$

$$\mathcal{F} = W_2\sigma(W_1 x)$$

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Global Average Pooling [2]

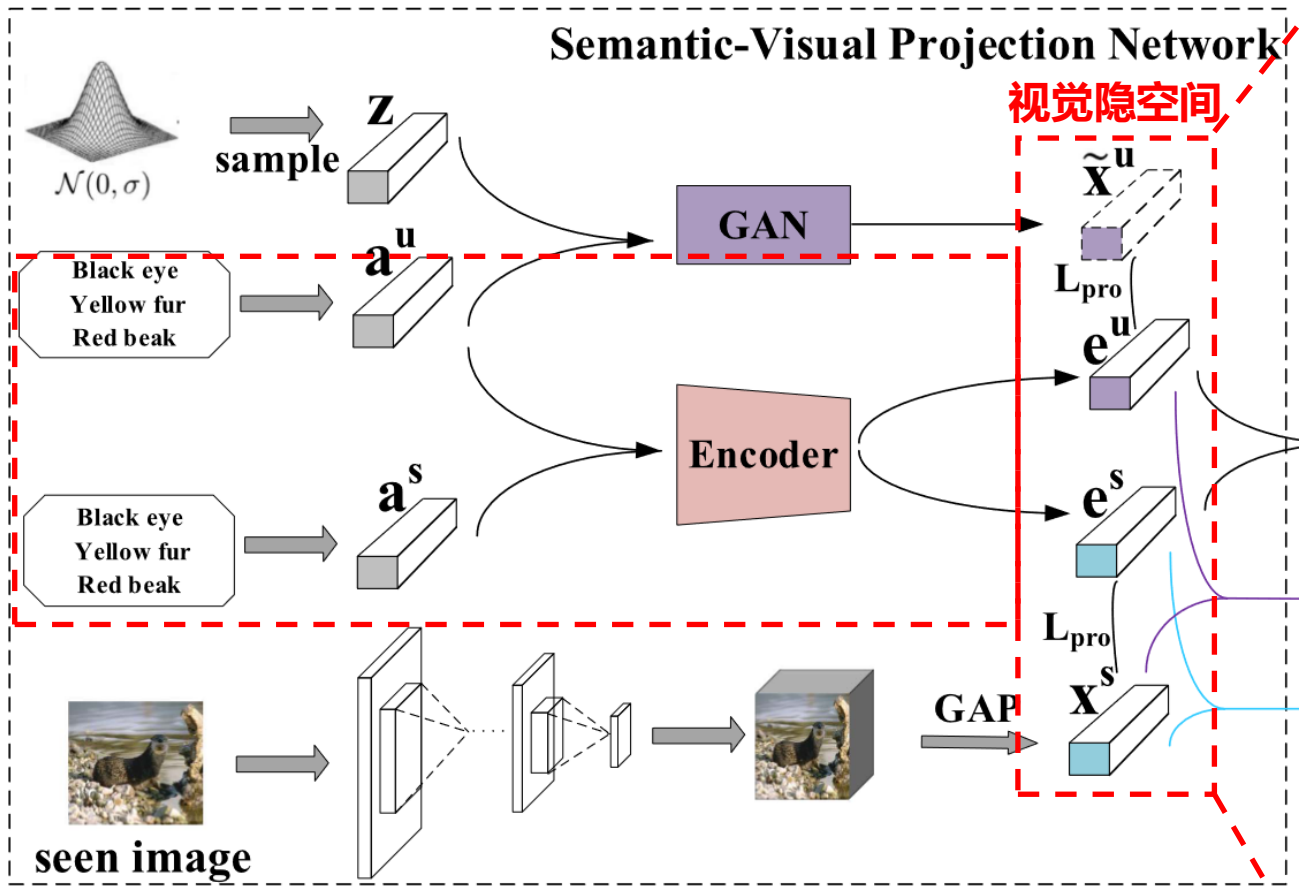


不可见类别视觉生成特征 $\tilde{X}^u$
可见类别视觉特征表示 X^s

监督训练 → 语义-视觉投影编码器

[1] He K, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016
[2] Shustanov A V, et al. Modification of single-purpose CNN for creating multi-purpose CNN[C]//Journal of Physics: Conference Series. IOP Publishing, 2019.

研究方法 | 2. 语义-视觉复合投影网络



2.3 语义-视觉投影

语义-视觉投影网络Encoder

Leaky ReLU Leaky ReLU

$$f_e(\mathbf{a}) = \delta_2(\mathbf{W}_{e2}\delta_1(\mathbf{W}_{e1}\mathbf{a} + \mathbf{b}_{e1}) + \mathbf{b}_{e2})$$

$$\mathbf{e}^s = f_e(\mathbf{a}^s) \quad \mathbf{e}^u = f_e(\mathbf{a}^u)$$

投影损失

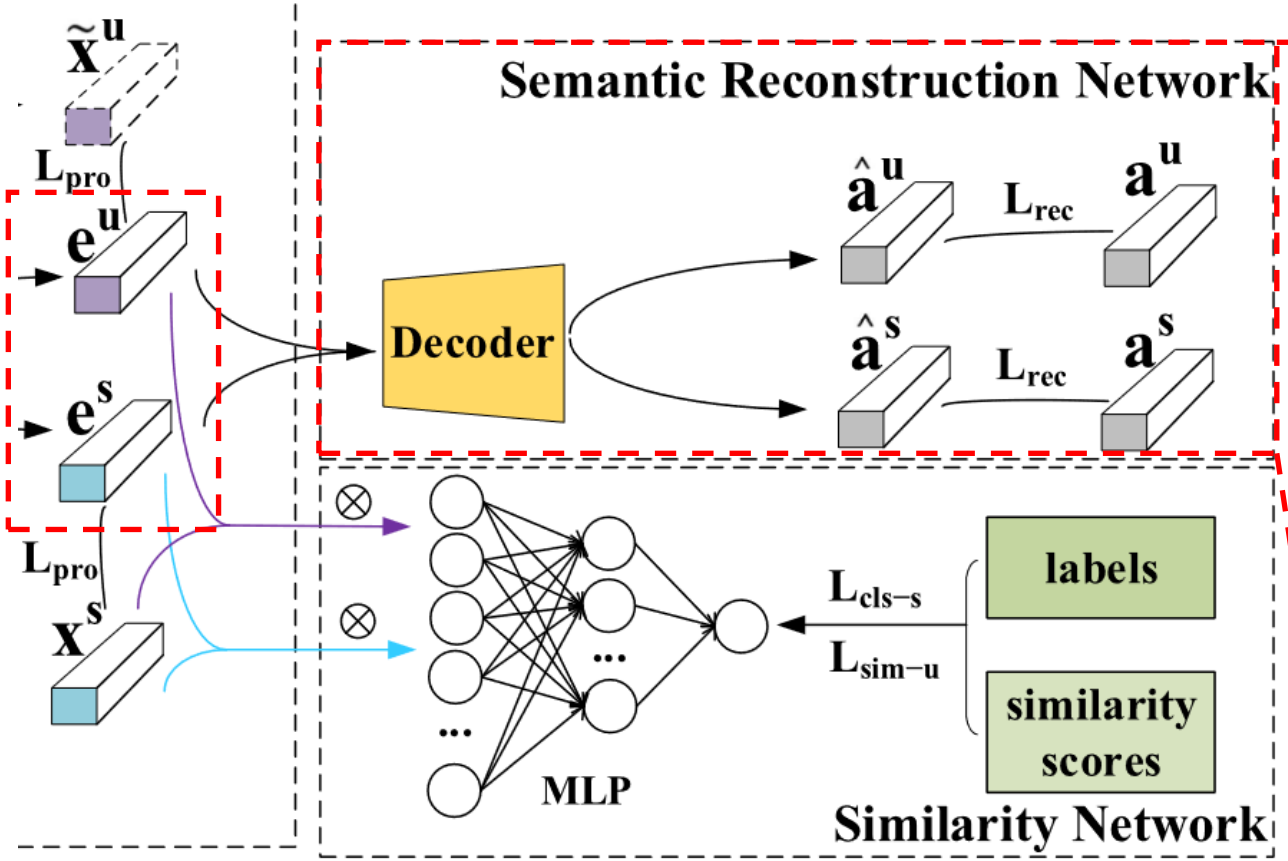
$$\mathcal{L}_{pro} = \underbrace{\frac{1}{N_s}\sum_{i=1}^{N_s}||\mathbf{e}_i^s - \mathbf{x}_i^s||_2^2}_{\text{Term 1}} + \underbrace{\frac{1}{\hat{N}_u}\sum_{i=1}^{\hat{N}_u}||\mathbf{e}_i^u - \tilde{\mathbf{x}}_i^u||_2^2}_{\text{Term 2}}$$

Term1: 对齐**可见类别**样本**语义与视觉特征**

Term2: 对齐**不可见类别**样本**语义与生成视觉特征**

视觉空间作为隐空间：探究语义和视觉特征之间映射关系
构建投影损失：对齐语义与视觉特征
使用生成样本训练：减少可见类别样本的偏差

研究方法 | 3. 语义重建网络



语义重建网络Decoder

Sigmoid ReLU

$$f_d(\mathbf{x}) = \delta_2(\mathbf{W}_{d2}\delta_1(\mathbf{W}_{d1}\mathbf{x} + \mathbf{b}_{d1}) + \mathbf{b}_{d2})$$

$$\hat{\mathbf{a}}^s = f_d(\mathbf{e}^s) \quad \hat{\mathbf{a}}^u = f_d(\mathbf{e}^u)$$

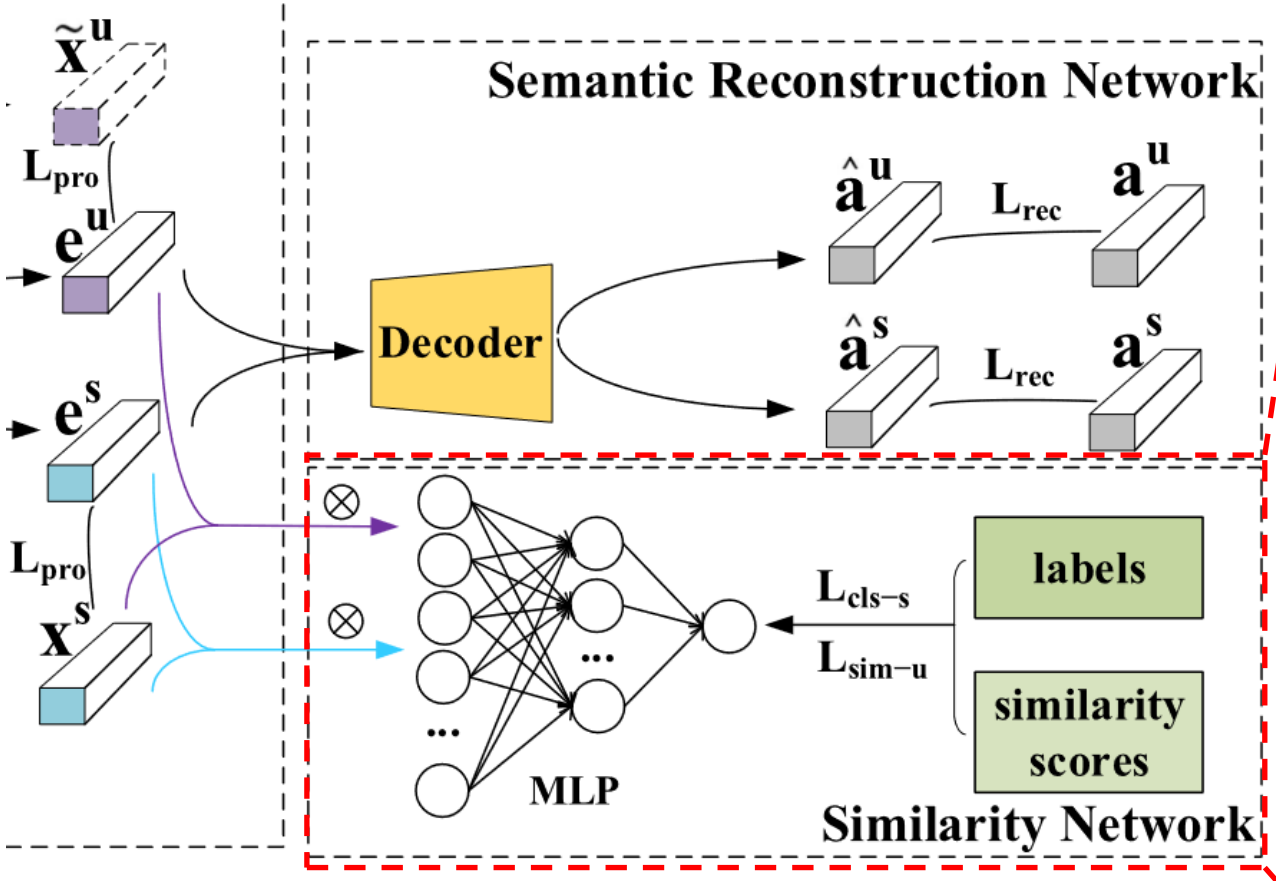
重建损失

$$\mathcal{L}_{rec} = \frac{1}{N_s}\sum_{i=1}^{N_s}||\hat{\mathbf{a}}_i^s - \mathbf{a}_i^s||_2^2 + \frac{1}{\hat{N}_u}\sum_{i=1}^{\hat{N}_u}||\hat{\mathbf{a}}_i^u - \mathbf{a}_i^u||_2^2$$

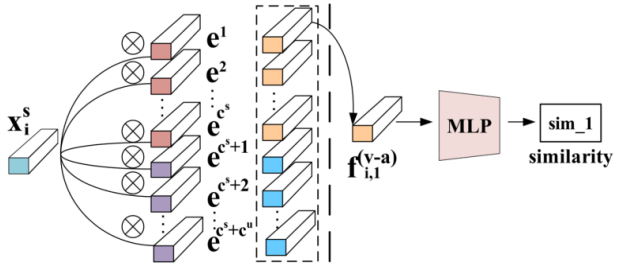
1. Decoder输入seen、unseen隐空间特征
2. 输出对应的原始语义特征
3. 通过重建损失进行训练

构建Decoder网络基于隐层特征重建原始语义特征：使隐空间特征保持语义辨别力

研究方法 | 4. 相似性网络



4.1 可见类别的分类损失



特征融合 $f_i^{(v-a)} = [\mathbf{x}_i^s \otimes \mathbf{e}^1, \mathbf{x}_i^s \otimes \mathbf{e}^2, \ldots, \mathbf{x}_i^s \otimes \mathbf{e}^{c^s+c^u}]$

Sigmoid Leaky ReLU

MLP $z_{ik} = \delta_2(\mathbf{W}_{s2}\delta_1(\mathbf{W}_{s1}f_{i,k}^{(v-a)} + \mathbf{b}_{s1}) + \mathbf{b}_{s2})$

$z_{i,k}$ 表示第i个样本和第k个可见类别之间的相似性

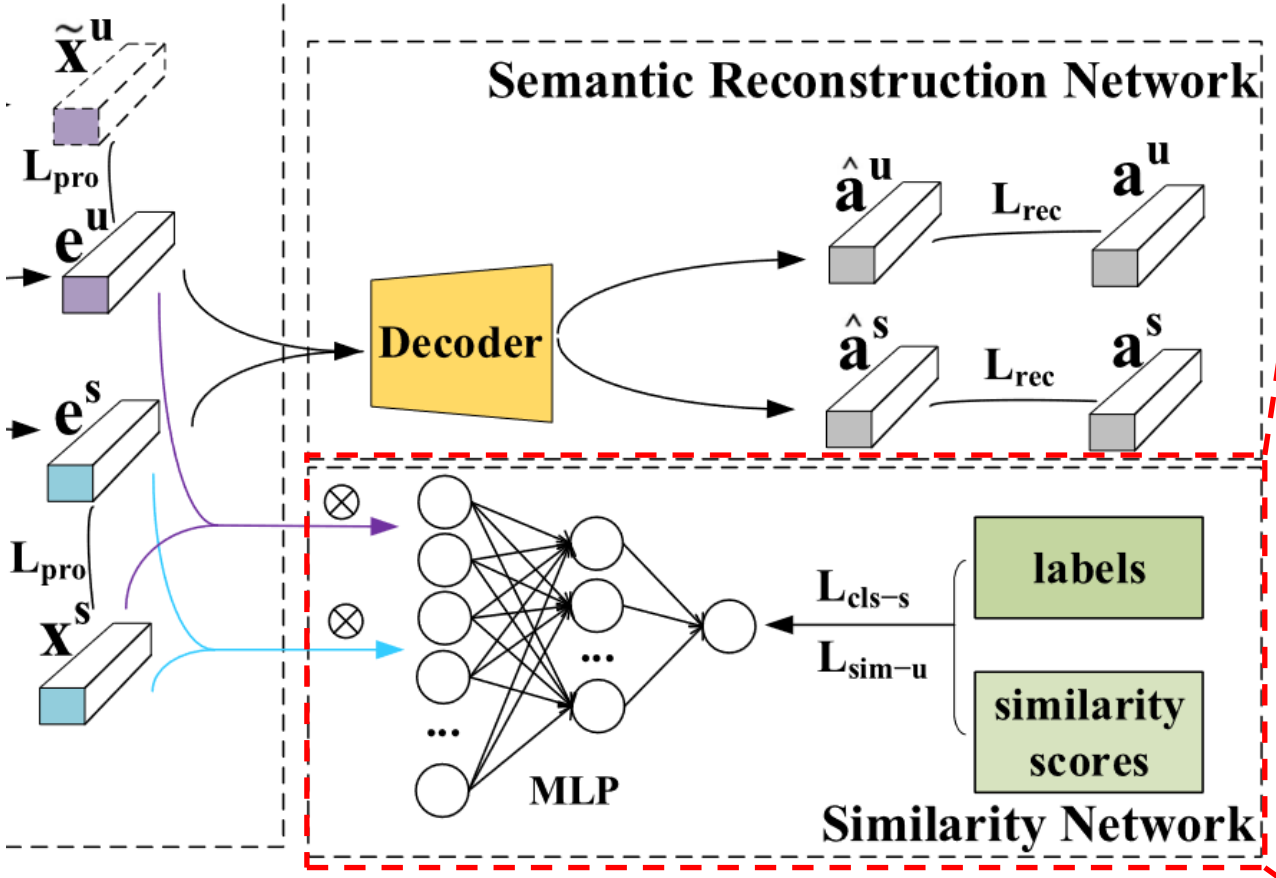
分类损失 $\mathcal{L}_{cls-s} = -\sum_{i=1}^{N_s}\sum_{k=1}^{c^s} m_{ik}logz_{ik} + (1-m_{ik})log(1-z_{ik})$

$m_{i,k}$ 第i个样本属于k可见类时值为1，否则值为0

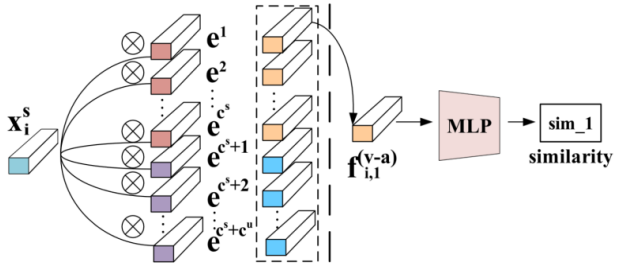
此时k的取值范围仅在可见类别数中 $k \leq c^s$

最小化分类损失：保证可见类别样本的视觉特征与其编码后的语义特征之间的相似性

研究方法 | 4. 相似性网络



4.2 未见类别的类间相似性损失



特征融合 $f_i^{(v-a)} = [\mathbf{x}_i^s \otimes \mathbf{e}^1, \mathbf{x}_i^s \otimes \mathbf{e}^2, \ldots, \mathbf{x}_i^s \otimes \mathbf{e}^{c^s+c^u}]$

Sigmoid Leaky ReLU

MLP $z_{ik} = \delta_2(\mathbf{W}_{s2}\delta_1(\mathbf{W}_{s1} f_{i,k}^{(v-a)} + \mathbf{b}_{s1}) + \mathbf{b}_{s2})$

$z_{i,k}$ 表示第i个样本和第k个未见类别之间的相似性

类间损失 $\mathcal{L}_{sim-u} = -\sum_{i=1}^{N_s} \sum_{k=c^s+1}^{c^s+c^u} p_{ik} log z_{ik} + (1-p_{ik}) log(1-z_{ik})$

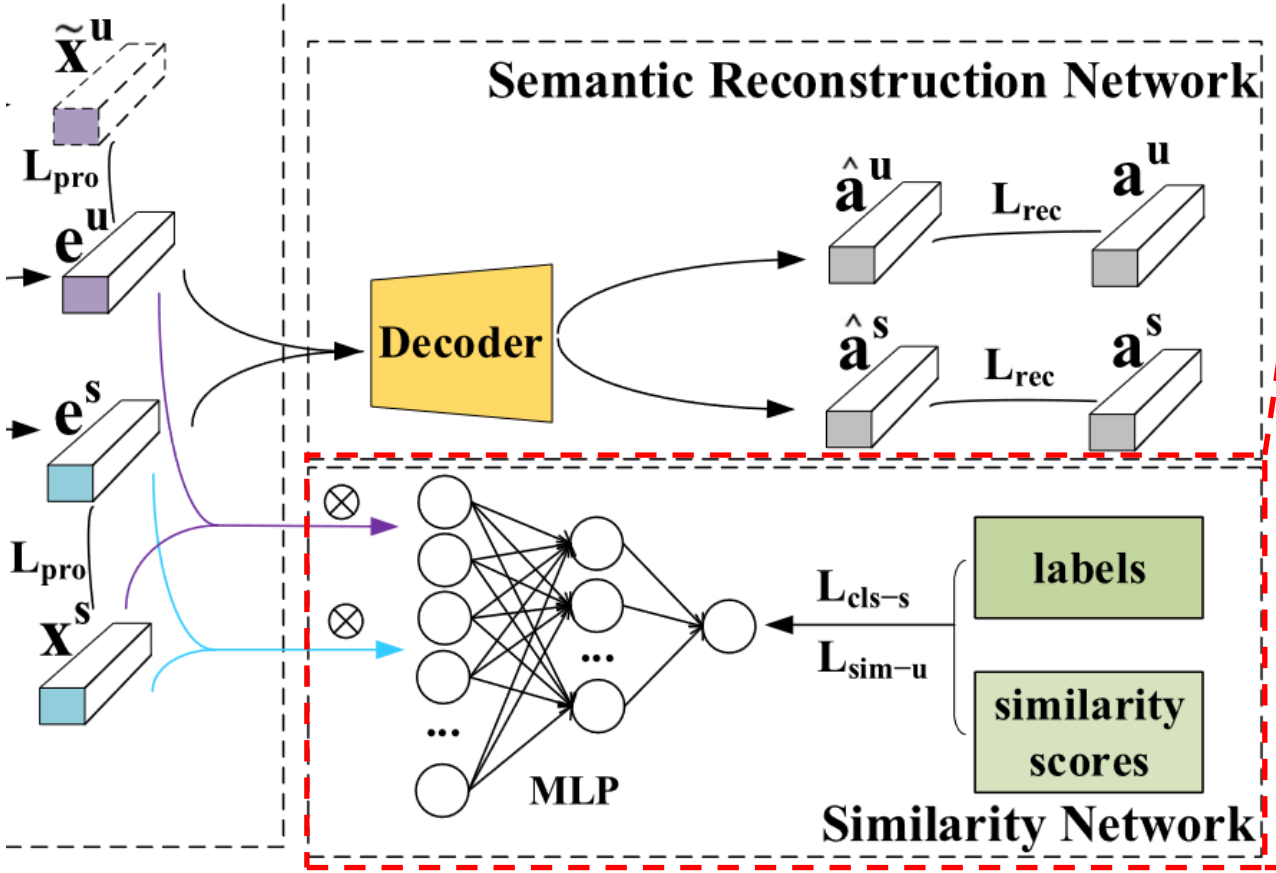
$p_{i,k}$ 表示第i个样本与第k个未见类别间的相似性分数

k的取值范围仅在未见类别数中 $c^s < k \leq c^s + c^u$

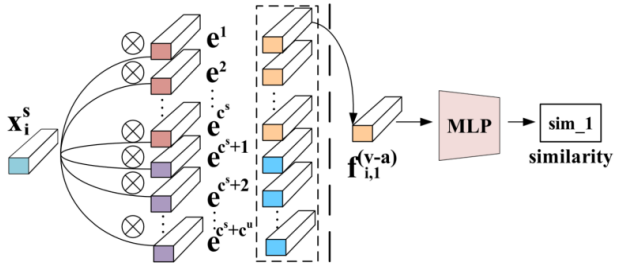
最小化类间损失：未见类别和最相关的可见类别具有最高相似性

挖掘了类间关系、将知识迁移给了未见类别

研究方法 | 4. 相似性网络



4.3 计算类间关系（软标签）



最小二乘回归 $\min_{\mathbf{P}} \|\mathbf{P}\mathbf{Z}^{u} - \mathbf{Z}^{s}\|_F^2 + \beta \|\mathbf{P}\|_F^2$

解析解 $\mathbf{P} = \mathbf{Z}^{s}\mathbf{Z}^{u\mathrm{T}}(\mathbf{Z}^{u}\mathbf{Z}^{u\mathrm{T}} + \beta\mathbf{I})^{-1}$

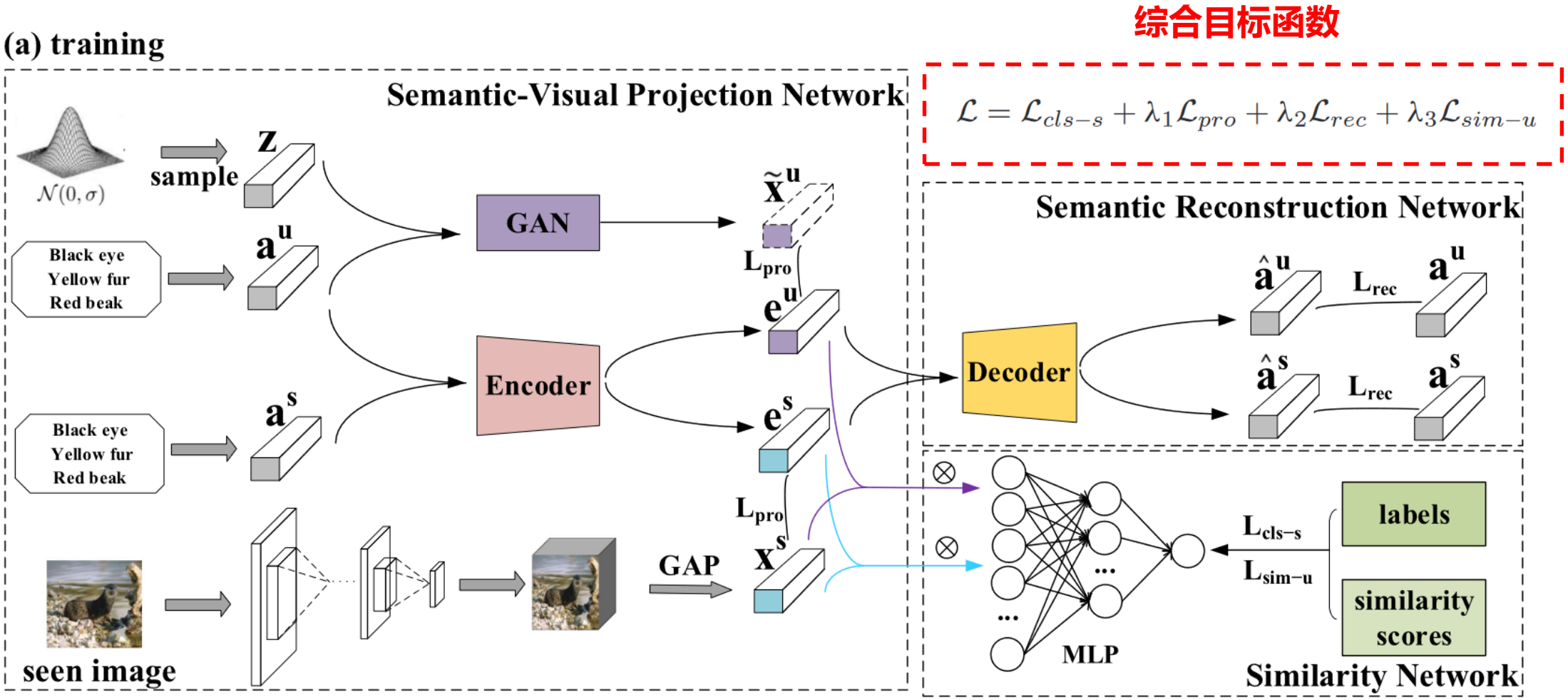
$\mathbf{Z}^{u}, \mathbf{Z}^{s}$ 分别表示未见类别和可见类别的原型语义表示

$\mathbf{P} \in \mathcal{R}^{c^{s} \times c^{u}}$ 表示类间关系，也是重建参数；第i行表示第i个可见类别与所有未见类别之间的类间相似度

语义流形：相似的视觉表示意味着相似的语义

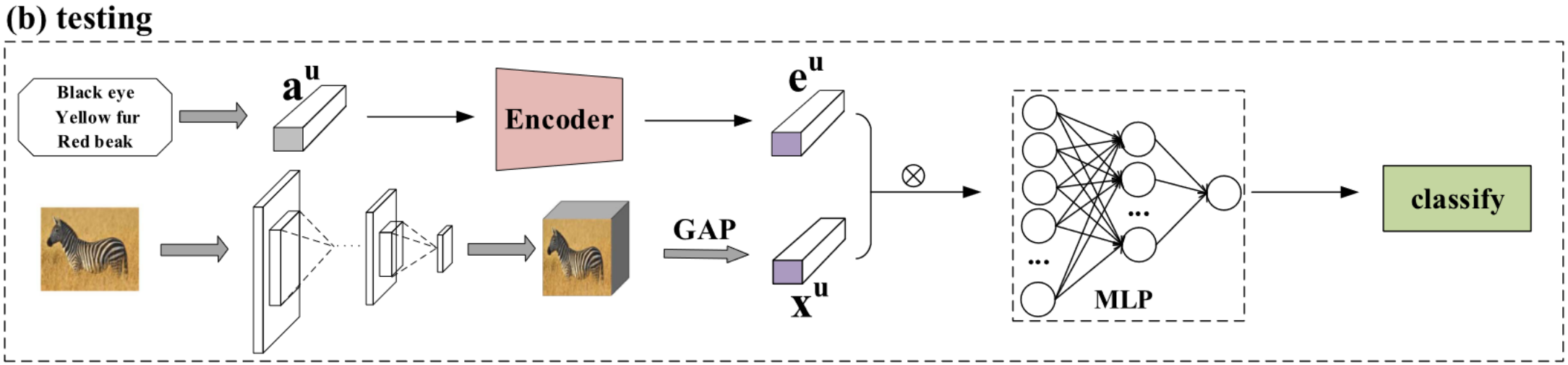
由可见类别标签和未见类别软标签训练：保证语义-视觉投影网络的辨别力、减小未见类别的语义混淆

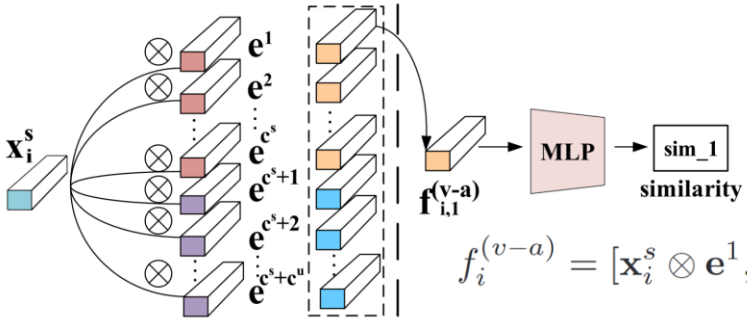
研究方法 | 5. 综合目标函数



基于上述的语义-视觉投影网络、语义重建网络、相似性网络构建模型，利用目标函数进行训练

研究方法 | 6. 零样本识别





$$f_i^{(v-a)} = [\mathbf{x}_i^s \otimes \mathbf{e}^1, \mathbf{x}_i^s \otimes \mathbf{e}^2, \ldots, \mathbf{x}_i^s \otimes \mathbf{e}^{c^s+c^u}]$$

$$z_{ik} = \delta_2(\mathbf{W}_{s2}\delta_1(\mathbf{W}_{s1}f_{i,k}^{(v-a)} + \mathbf{b}_{s1}) + \mathbf{b}_{s2})$$

零样本学习ZSL

$$\hat{y}(x_i) = \underset{j \in \mathcal{Y}^u}{\operatorname{argmax}}\, z_{ij}$$

广义零样本学习GZSL

$$\hat{y}(x_i) = \underset{j \in \mathcal{Y}^s \cup \mathcal{Y}^u}{\operatorname{argmax}}\, z_{ij}$$

$z_{i,j}$ 表示第i个样本和第j个类别之间的相似性

根据相似性进行分类，对应相似性最大的类别即是该样本的分类类别

目录

CONTENTS

实验分析 | 1. 实验设置

TABLE I
STATISTICS OF EACH DATASET. SS: SEMANTIC SPACE; 'A': ATTRIBUTE; 'W': WORD VECTOR; SS-DIM: DIMENSION OF SEMANTIC VECTOR; S/U: SEEN/UNSEEN CLASS SIZE; TOTAL: NUMBER OF ALL INSTANCES

Dataset	SS	SS-Dim	S/U	Total
AWA1	A	85	40/10	30475
AWA2	A	85	40/10	37322
CUB	A	312	150/50	11788
SUN	A	102	645/72	14340
APY	A	64	20/12	15339
FLO	W	1024	82/20	8189
ImageNet	W	1000	1000/360	254000

1.1 数据集

AWA1、AWA2：动物类别；**CUB**：细分鸟类；
SUN：场景类别；**APY**：小型分类数据集；
FLO：花类别；**ImageNet**：大型分类数据集

SS：语义空间；**SS-Dim**：语义向量维度；
S/U：可见/未见类别数；**Total**：样本总数；

A：类别的属性
W：词向量

1.2 数据集划分

注意：避免未见类别出现在预训练过程中，失去公平性

1.3 评估标准

T_u：未见类别的平均类别准确率

T_s：可见类别的平均类别准确率

调和平均值H

$$H = \frac{2 \times (T\_s \times T\_u)}{T\_s + T\_u}$$

1.4 实现细节- Pytorch、Adam

β：{0.001, 0.01, 0.1}　　**逐类别交叉验证策略**

$\lambda_1, \lambda_2, \lambda_3$: {0.0001, 0.001, 0.01, 0.1}

实验分析 | 2. 实验结果比较

2.1 与SOTA结果的比较 – 零样本学习ZSL

TABLE II
COMPARISON WITH THE STATE-OF-THE-ART METHODS UNDER ZSL SETTING

Method	AWA1	AWA2	CUB	SUN	APY	FLO
DEVISE[21]	54.2	59.7	52.0	56.5	39.8	45.9
ALE [17]	59.9	62.5	54.9	58.1	39.7	48.5
ESZSL [25]	58.2	58.6	53.9	54.5	38.3	51.0
SAE [47]	53.0	54.1	33.3	40.3	8.3	-
DEM[53]	68.4	67.1	51.7	40.3	35.0	-
SYNC[56]	54.0	46.6	55.6	56.3	23.9	-
RN[3]	68.2	64.2	55.6	-	-	-
f-CLSWGAN[36]	68.2	-	57.3	60.8	-	67.2
SE-ZSL[57]	69.5	69.2	59.6	-	-	-
PSR[58]	-	63.8	56.0	61.4	38.4	-
SP-AEN[48]	-	58.5	55.4	59.2	24.1	-
QFSL[59]	-	63.5	58.8	56.2	-	-
GAZSL[60]	-	68.2	55.8	61.3	-	60.5
DCN[61]	-	65.2	56.2	61.8	43.6	-
ALS[62]	-	66.2	57.5	62.0	**44.5**	-
MGA-GAN[63]	70.6	-	58.9	61.8	-	**70.6**
Ours	**70.9**	**70.4**	**61.0**	**62.2**	39.9	62.7

在ZSL数据集上取得了卓越的性能表现

优点：

1. AWA1、AWA2、CUB、SUN上取得最好的性能
2. CUB、SUN上证明CPL能更好地分类细粒度目标
3. CPL参数更少，更易于训练

缺点：

1. APY上性能略差，类间关系弱
2. FLO上性能较差，缺少判别力

实验分析 | 2. 实验结果比较

2.2 与SOTA结果的比较 – 广义零样本学习GZSL & 大型数据集ZSL

TABLE III
COMPARISON WITH THE STATE-OF-THE-ART METHODS UNDER GZSL SETTING ON ZSL DATASETS. WE EMPLOY HARMONIC MEAN H TO MEASURE THE CLASSIFICATION RESULTS

	Generalized Zero-Shot Learning																	
	AWA1			AWA2			CUB			SUN			APY			FLO		
Method	T_u	T_s	H	T_u	T_s	H	T_u	T_s	H	T_u	T_s	H	T_u	T_s	H	T_u	T_s	H
DEVISE[21]	13.4	68.7	22.4	17.1	74.7	27.8	23.8	53.0	32.8	16.9	27.4	20.9	4.9	76.9	9.2	9.9	44.2	16.2
ALE [17]	16.8	76.1	27.5	14.0	81.8	23.9	23.7	62.8	34.4	21.8	33.1	26.3	4.6	73.7	8.7	13.3	61.6	21.9
ESZSL [25]	6.6	75.6	12.1	5.9	77.8	11.0	12.6	63.8	21.0	11.0	27.9	15.8	2.4	70.1	4.6	11.4	56.8	19.0
SAE [47]	1.8	77.1	3.5	1.1	82.2	2.2	7.8	54.0	13.6	8.8	18.0	11.8	0.4	**80.9**	0.9	-	-	-
PSR[58]	-	-	-	20.7	73.8	32.3	24.6	54.3	33.9	20.8	37.2	26.7	13.5	51.4	21.4	-	-	-
QFSL[59]	-	-	-	52.1	72.8	60.7	33.3	48.1	39.4	30.9	18.5	23.1	-	-	-	-	-	-
SE-ZSL[57]	56.3	67.8	61.5	58.3	68.1	62.8	41.5	53.3	46.7	-	-	-	-	-	-	-	-	-
LFGAA[64]	-	-	-	27.0	**93.4**	41.9	36.2	**80.9**	50.0	18.5	40.0	25.3	-	-	-	-	-	-
DEM[53]	32.8	**84.7**	47.3	30.5	86.4	45.1	19.6	57.9	29.2	20.5	34.3	25.6	11.1	75.1	19.4	-	-	-
GAZSL[60]	-	-	-	19.2	86.5	31.4	23.9	60.6	34.3	21.7	34.5	26.7	-	-	-	28.1	77.4	41.2
TCN [35]	49.4	76.5	60.0	**61.2**	65.8	63.4	52.6	52.0	52.3	31.2	37.3	34.0	24.1	64.0	35.1	-	-	-
CRNet[65]	58.1	74.7	65.4	52.6	78.8	63.1	45.5	56.8	50.5	34.1	36.5	35.3	32.4	68.4	44.0	-	-	-
APNet[66]	59.7	76.6	**67.1**	54.8	83.9	66.4	48.1	55.9	51.7	35.4	**40.6**	37.8	32.7	74.7	45.5	-	-	-
ZSKL[67]	17.9	82.2	29.4	18.9	82.7	30.8	21.6	52.8	30.6	20.1	31.4	24.5	10.5	76.2	18.5	-	-	-
MSEL[68]	52.6	76.7	62.4	52.3	81.3	63.7	-	-	-	-	-	-	28.4	75.5	41.2	56.1	81.2	66.4
DASCN[69]	59.3	68.0	63.4	-	-	-	45.9	59.0	51.6	42.4	38.5	40.3	**39.7**	59.5	**47.6**	-	-	-
RFF(softmax)[70]	**59.8**	75.1	66.5	-	-	-	52.6	56.6	54.6	**45.7**	38.6	**41.9**	-	-	-	**65.2**	78.2	**71.1**
MLSE[71]	-	-	-	23.8	83.2	37.0	22.3	71.6	34.0	52.3	24.3	33.2	-	-	-	-	-	-
DAZLE[72]	-	-	-	60.3	75.7	67.1	56.7	59.6	**58.1**	20.7	36.4	26.4	12.7	74.3	21.7	-	-	-
ZSML[73]	57.4	71.1	63.5	58.9	74.6	65.8	**60.0**	52.1	55.7	-	-	-	36.3	46.6	40.9	-	-	-
ALS[62]	-	-	-	53.8	56.0	54.9	43.1	51.6	46.9	41.5	31.9	36.1	28.6	65.5	40.0	-	-	-
MGA-GAN[63]	59.3	67.7	63.2	-	-	-	46.6	58.3	51.8	45.6	37.3	41.0	-	-	-	58.4	**81.9**	68.2
Ours	59.1	73.9	65.7	60.0	76.8	**67.4**	46.5	59.4	52.2	45.5	33.3	38.5	24.1	60.9	34.5	51.8	69.4	59.3

TABLE IV
COMPARISON WITH THE STATE-OF-THE-ART METHODS UNDER ZSL SETTING ON IMAGENET. THE BEST RESULTS ARE IN **BOLD**

Method	Generative	ZSL
CONSE[22]	✗	7.8
DEVISE[21]	✗	5.2
AMP[76]	✗	6.1
SS-Voc[77]	✗	9.5
Ours	✓	**10.0**

分析

1. T_u指标上有竞争力的结果，bias少
2. AWA2上取得最好的H值，balance好
3. 结构比其他基于生成的方法简单
4. RFF性能更好，但结构更复杂
5. ImageNet平均准确率取得最好的结果

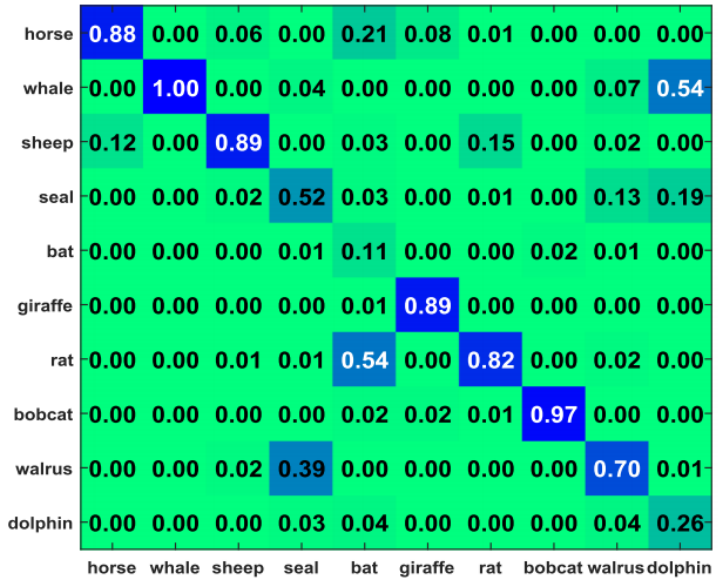
在各种不同数据集上取得了较好的性能表现

实验分析 | 3. 实验分析与讨论

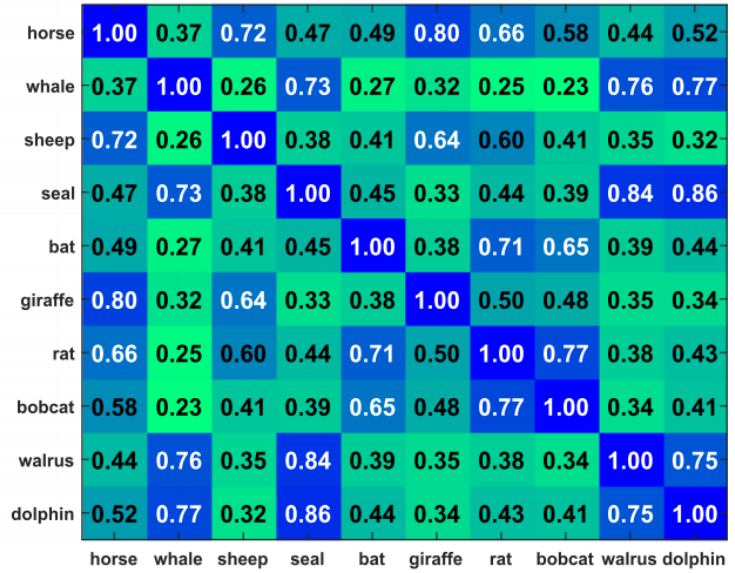
3.1 模型分析与讨论

AWA2数据集上ZSL

分类准确率混淆矩阵



语义余弦相似度



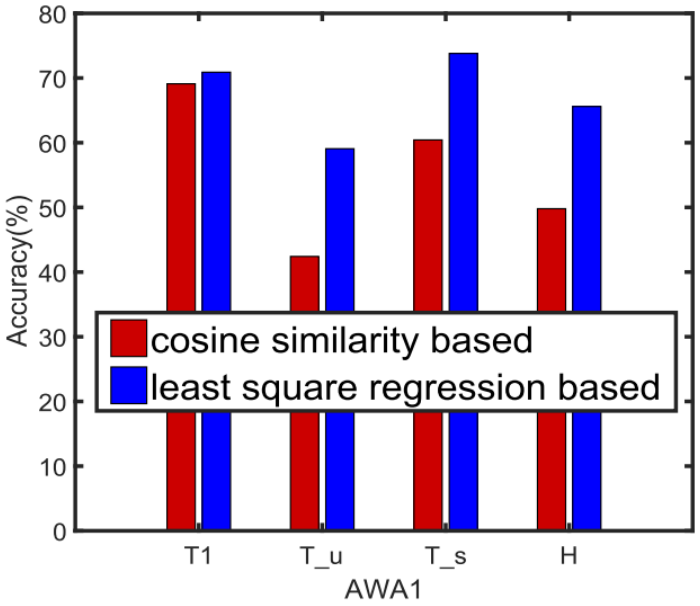
1. 大多超70%准确率，部分超90%准确率
2. bat和dolphin准确率较差
3. 相似度越高，分类概率越大
4. 混淆矩阵刻画了语义相似度

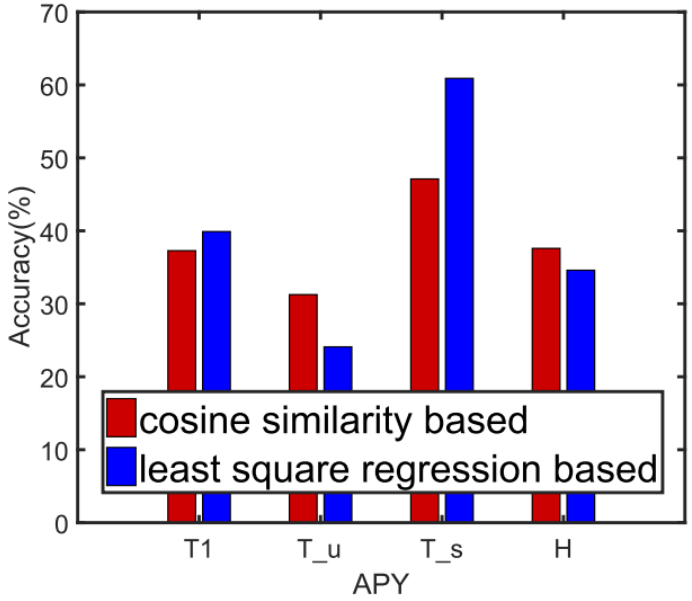
该实验证明了提出方法在零样本识别任务的有效性和优越性

实验分析 | 3. 实验分析与讨论

3.1 模型分析与讨论

不同建模类间关系方法对分类性能的影响





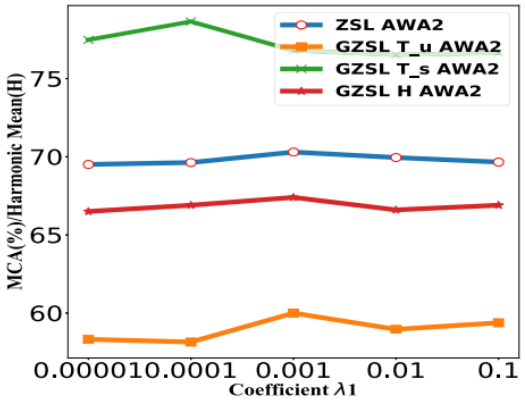
1. AWA1上最小二乘回归比余弦相似度好
2. APY上余弦相似的调和均值H更大
3. 综合考虑ZSL和GZSL，最小二乘更优

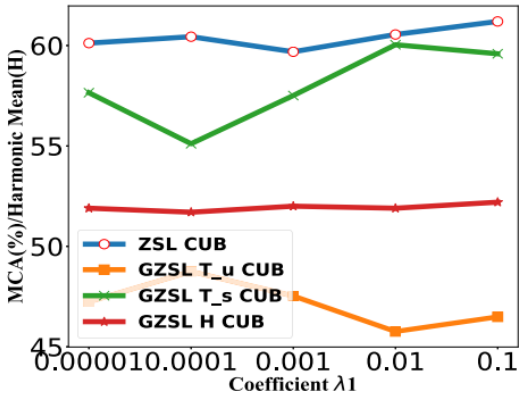
综上，采用基于最小二乘的方法来计算类间相似度

实验分析 | 3. 实验分析与讨论

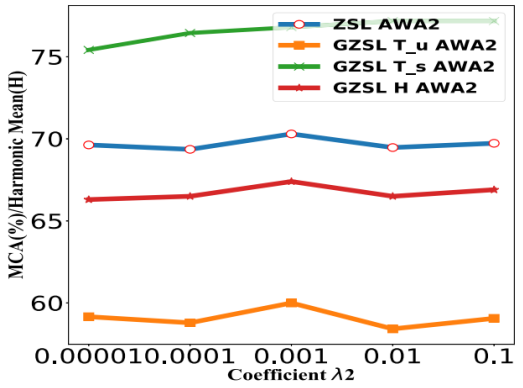
3.2 参数敏感度 - 先搜寻最优配置，后观察各参数对性能的影响

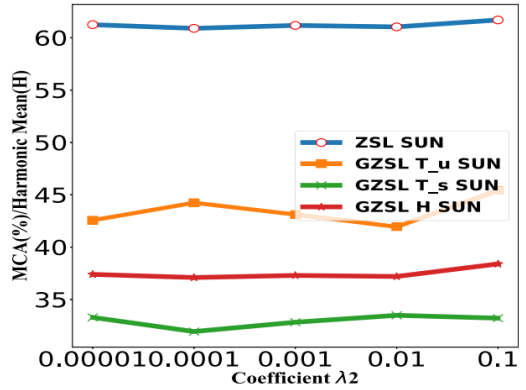
参数λ_1对ZSL和GZSL性能的影响





参数λ_2对ZSL和GZSL性能的影响





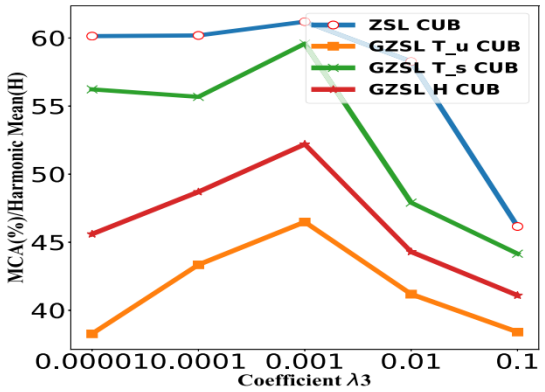
1. 调整该参数**对性能影响不大**
2. 调和均值**H的数值相对稳定**
3. **平衡可见和未见类别准确率能够取得更好的泛化性能**

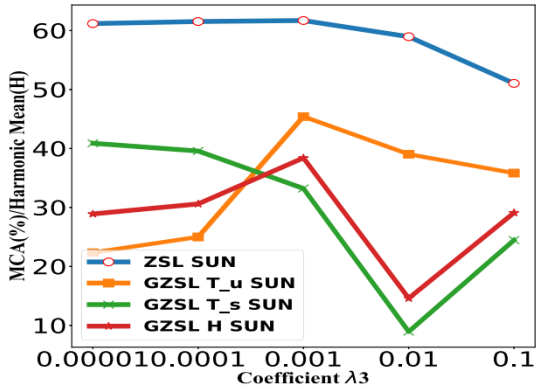
1. 两个数据集对该参数**比较鲁棒**
2. SUN数据集上**准确率更稳定**

实验分析 | 3. 实验分析与讨论

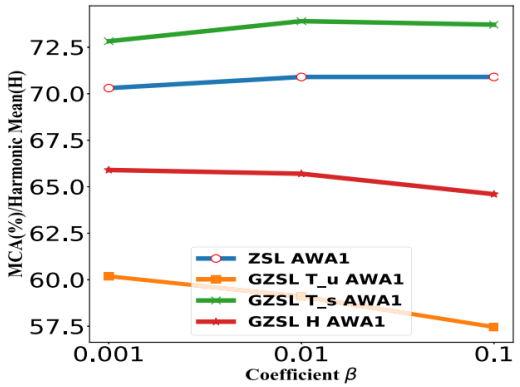
3.2 参数敏感度 - 先搜寻最优配置，后观察各参数对性能的影响

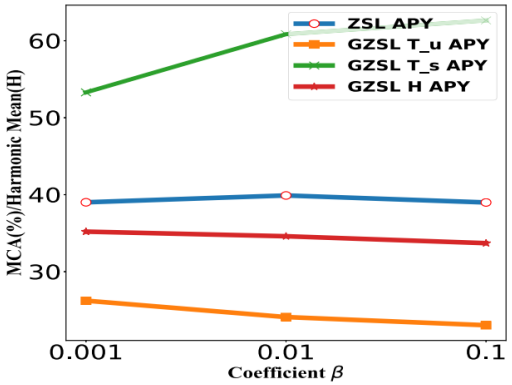
参数λ_3对ZSL和GZSL性能的影响





参数β对ZSL和GZSL性能的影响





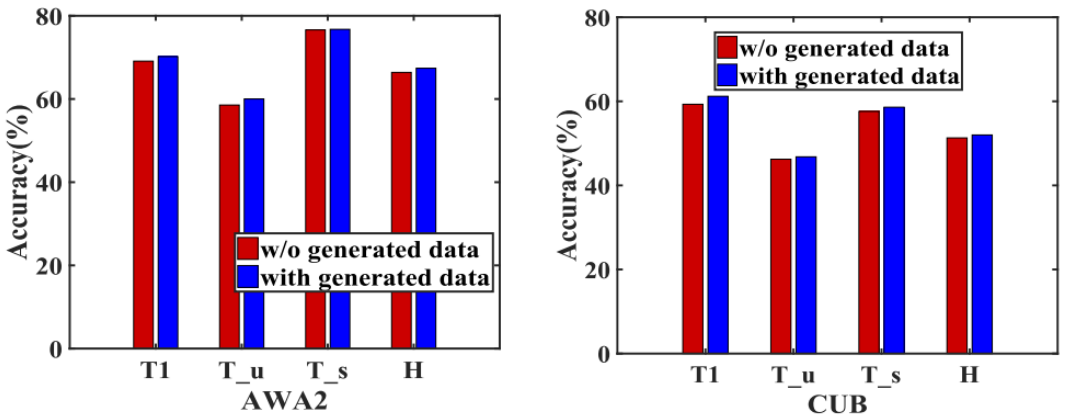
1. 该参数数值小于0.001时ZSL性能较好
2. 该参数等于0.001时GZSL的H值最高
3. 提出的方法对该参数敏感，需要合适地调节

该参数对两个数据集都比较鲁棒

实验分析 | 3. 实验分析与讨论

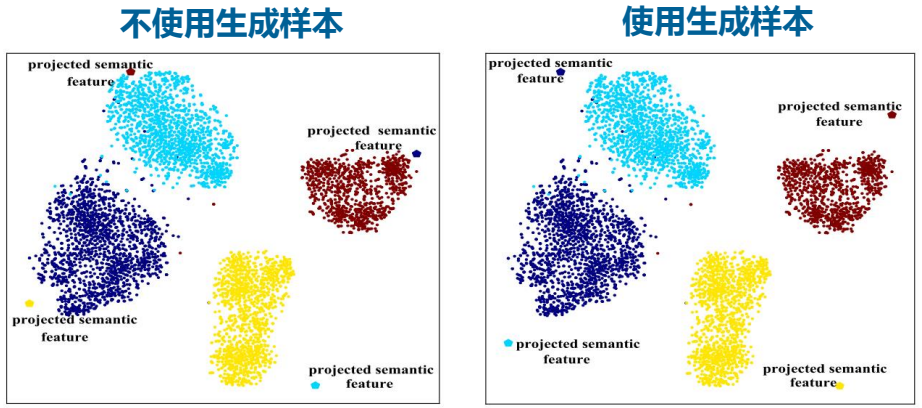
3.3 消融试验

生成未见类别样本的影响



1. 引入生成样本**提升了未见类别准确率T_u**
2. 引入生成样本**提升了调和均值H**
3. **证明该方法帮助减轻了偏移问题**

未见类别样本的可视化

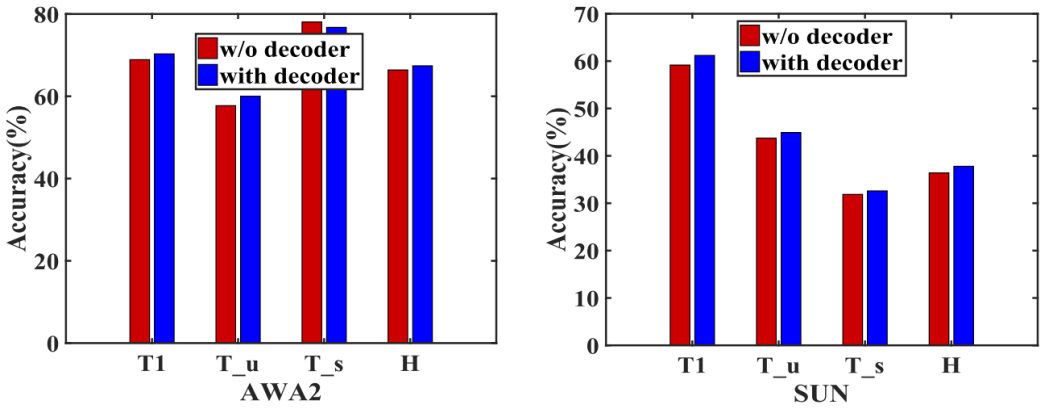


1. 视觉特征具有**更好的分类特性**
2. **编码的语义特征与其视觉特征更加接近**
3. 说明视觉与语义两个**模态对齐较好，且偏差问题缓解**

实验分析 | 3. 实验分析与讨论

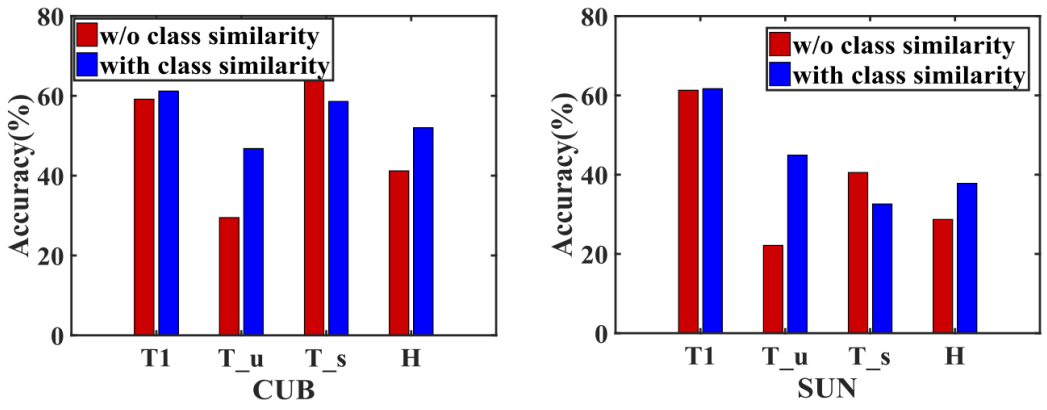
3.3 消融试验

语义重建网络的影响



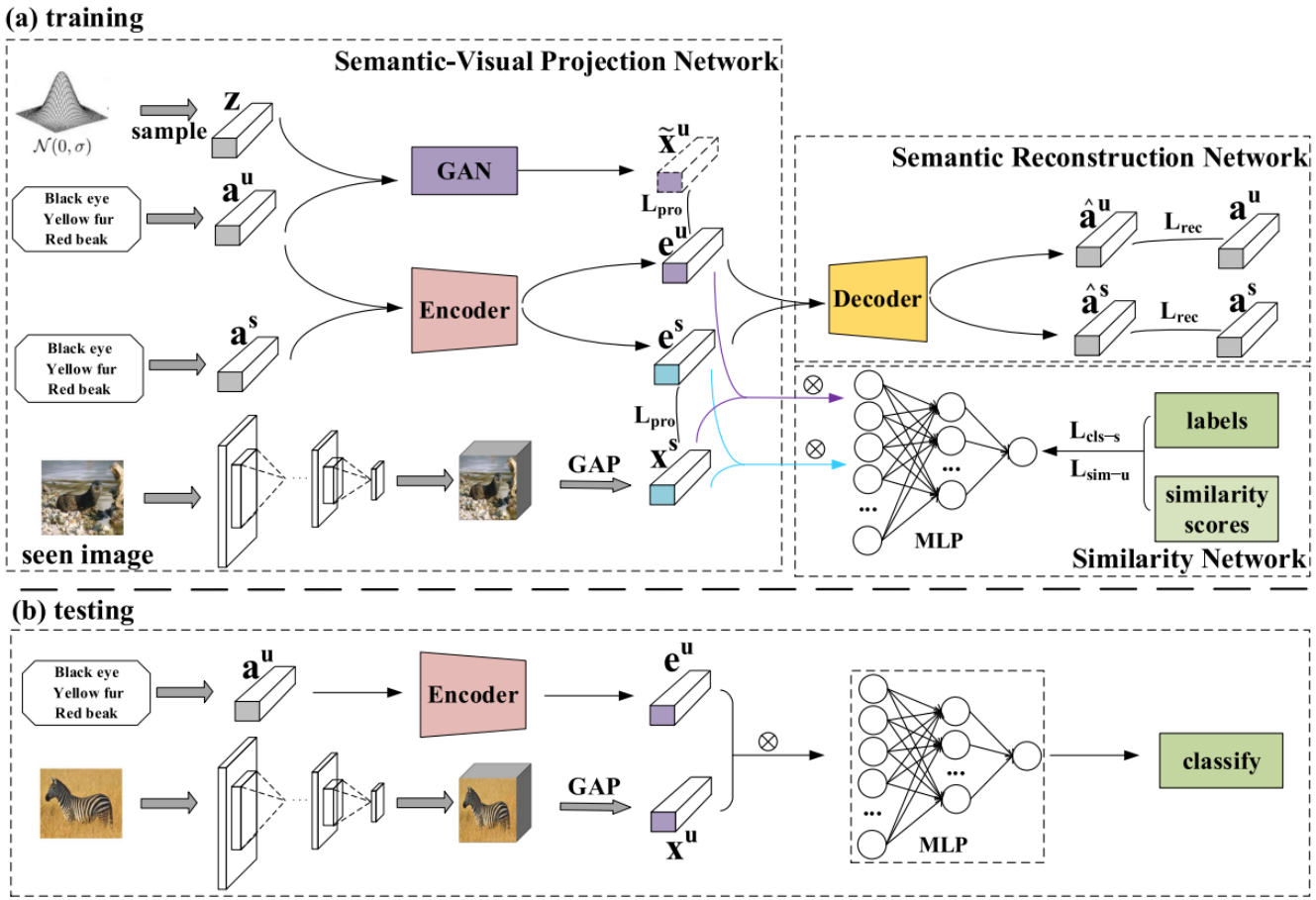
1. 重建网络帮助性能提升
2. 准确利用视觉空间的低层丰富特征信息

类间相似性的影响



1. 对ZSL影响较小，对GZSL影响较大
2. 没有类间相似性，H值较低
3. 证明该方法帮助可见样本和未见样本的性能平衡

总结与展望



1. 本文提出复合投影学习CPL用于零样本学习任务
2. 利用可见样本和未见生成样本减少偏差问题
3. 相似性网络改善了编码语义特征的混淆问题
4. 实验结果证明了提出方法的优越性
5. 未来考虑利用迁移学习来用于ZSL任务

感谢聆听