基于组间协作学习的协同显著目标检测

范琦 [1,3,] 范登平 [2,*] 付华柱 [2] Chi-Keung Tang[1] 邵岭 [2] Yu-Wing Tai[1,3]
[1] 香港科技大学 [2] 起源人工智能研究院 (IIAI) [3] 快手科技

摘要

本文提出了一种新的组间协作学习框架 *(**GCoNet**)*，该框架能够实时 *(16ms)* 地检测协同显著对象，基于两个必要条件同时挖掘组级共识 *(consensus)* 表示：***1)* 组内紧凑性 *(intra-group compactness)***，通过使用本文有创意的组内亲和力模块来捕捉协同显著对象的内在共享属性，更好地表达协同显著对象之间的一致性；***2)* 组间可分性 *(inter-group separability)***，通过引入一种基于差异共识的组间协作模块来有效地抑制噪声对象对输出的影响。为了在不增加额外计算开销的情况下学习更好的嵌入空间，本文显式地使用了辅助分类监督。在三个具有挑战性的基准数据集 *(即 CoCA、CoSOD3k 和 Cosal2015)* 上的广泛实验表明，本文简单的 *GCoNet* 超越了 *10* 个尖端模型，并达到了新的最好性能。本文介绍了新技术在一些重要的下游计算机视觉应用上的贡献，包括内容感知的协同分割、基于协同定位的自动缩略图等。代码地址为：*https://github.com/fanq15/GCoNet*。

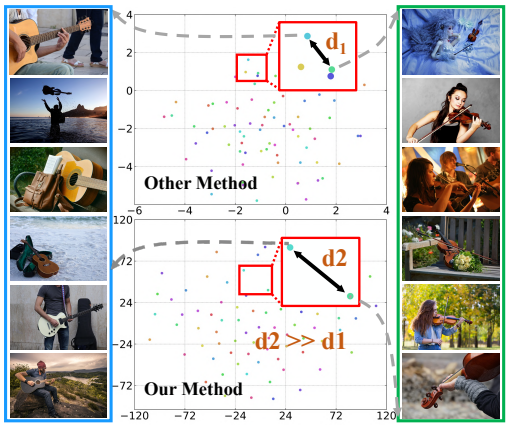

图 1. t-SNE [19] 可视化共识，其中每个点代表图像组的一个共识。这里采用了两个相似但不同的组 (吉他 & 小提琴) 来证明本文的 GCoNet 的有效性。传统 CoSOD 模型 (CoEGNet [9]) 中的共识策略倾向于将共识聚在一起，即使它们属于不同的组，导致协同显著检测产生歧义。而本文的共识策略具备有效的组间约束，能够增大不同组间的距离 ($d_2 \gg d_1$)，从而实现更好的组间可分性。

1. 引言

在给定一组相关图像的情况下，协同显著对象检测 (Co-salient object detection CoSOD) 用于检测具有相同属性的协同显著对象。CoSOD 比标准的显著对象检测 (SOD) [2, 3, 4] 任务和基于深度图的显著目标检测 (RGB-D SOD) [5, 6, 7, 8] 更具有挑战性，因为 CoSOD 需要在存在其他对象 (干扰对象) 的情况下区分多个图像 [9] 上共同出现的对象。也就是说，应该同时最大化类内紧凑性和类间可分性。因此，CoSOD 经常被用作各种计算机视觉任务的预处理步骤，例如图像检索 [10]、图像质量评估 [11]、基于集合的图片裁剪 [12]、协同分割 [13, 14]、语义分割 [15]、图像监视 [16]、视频分析 [17]、视频协同定位 [18] 等。

以前的工作试图利用相关图像之间的一致性，通过探索在一个图像组内不同的共享线索 [20, 21, 22] 或语义连接 [23, 24, 25] 来提高模型在 CoSOD 任务上的表现。[26, 27] 通过计算各种组内图像间的线索来获得显著图，从而发现协同显著对象。其他工作 [9, 28] 探索了一个统一网络来联合优化协同显著信息和显著图。

尽管其他方法取得了较好的效果，但目前的大多数模型都只在单个图像组中提取其 CoSOD 的特征表

*通讯作者：范登平 *(dengpfan@gmail.com)*。
[0]本文为 CVPR21 [1] 中译版，由邢浩哲译，范琦、范登平校稿。

示，这带来了一些限制。首先，同一组图像包含相似的前景 (即协同显著的对象)，只提供正向关系，而忽略不同对象之间的负向关系。仅使用正向关系训练模型可能会导致过拟合，并且困难图像的结果较差。此外，组中的图像数量通常是有限的 (对于大多数 CoSOD 数据集来说是 20 到 40 个图像)，因此使用单个组不能为学习有辨别力的特征提供足够的信息。最后，单个图像组也不能提供在复杂的现实场景中进行推理时区分噪声对象所必需的高层次的语义信息。

针对上述问题，本文提出了一种新的组间协作学习框架 (**GCoNet**) 来挖掘不同图像组之间的语义相关性。所提 GCoNet 由三个重要组成部分组成：组内亲和模块 (group affinity module GAM)、组间协作模块 (group collaborating module GCM) 和辅助分类模块 (auxiliary classification module ACM)，它们同时学习**组内紧凑性**和**组间可分性**。GAM 使网络学习同一图像组内的共识特征，而 GCM 区分不同组之间的目标属性，从而使模型能够在现有的大规模 SOD 数据集[1]上进行训练。除此之外，本文用 ACM 来进一步改进每幅图像的全局语义级的特征表示，以学习更好的特征空间。总而言之，本文的贡献是：

- 本文提出了一种新的组间协作学习策略来解决 CoSOD 问题，并通过大量的消融实验验证了该策略的有效性。
- 本文通过同时考虑组内紧凑性和组间可分性来挖掘共识特征，设计了一种新的用于 CoSOD 的统一组间协作学习网络。
- 本文的组内亲和模块 (GAM) 和组间协作模块 (GCM) 相互协作，以实现更好的组内和组间协作学习。辅助分类模块 (ACM) 进一步促进全局语义级的学习。
- 在 CoCA、CoSOD3k 和 Cosal2015 这三个具有挑战性的 CoSOD 基准 (benchmark) 数据集上的广泛实验表明，本文的 GCoNet 达到了最好性能。此外，基于本文的技术贡献，本文提供了两种下游应用，即协同分割和协同定位。

[1]现有的 CoSOD 数据集总共包含约 6 千张图像，而 SOD 数据集有 12 个以上，包含约 6 万张图像。这些 SOD 数据集可以缓解协同显著性物体检测中训练数据不足的问题。

2. 相关工作

传统的显著目标检测任务 [29, 30, 31, 32, 33] 的目标是分别直接分割每幅图像中的显著目标，而 CoSOD 的目标是在多幅相关图像中分割相同的显著目标。以往的工作主要是利用图像间的信息来检测协同显著的对象。早期的 CoSOD 方法基于人工设计的浅层特征 [18, 34]，来探索图像对 [20, 35] 或一组相关图像之间的物体对应关系。他们使用基于约束或启发式特征的方法来挖掘图像间的关系。一些研究试图通过使用高效的流形排序方案 (manifold ranking scheme) [36] 来获得具有引导效果的显著图，或使用全局关联约束与聚类 (global association constraint with clustering) [21] 或平移对齐 (translational alignment) [12] 来捕获图像间约束。其他工作试图从启发式特征中的高层特征、使用多显著线索 (multiple saliency cues) 和自适应权重 (self-adaptive weights) [22]、区域直方图和约束 (regional histograms and constrasts) [37]、基于优化新目标函数的度量学习 (metric learning by optimizing a new objective function) [24] 或成对相似性排序和线性规划 (pairwise similarity ranking and linear programming) [38] 来构建组中图像之间共享的语义属性。

最近，基于深度学习的模型通过不同的方法，如图卷积网络 (GCN) [39, 40, 41]、自学习 (self-learning) 方法 [23, 42]、基于 PCA 投影的图像间协同注意 (inter-image co-attention with PCA projection) [9] 或循环单元 (recurrent units) [9]、协同相关技术 (correlation techniques) [43]、质量度量 (quality measurement) [44] 或协同聚类 (co-clustering) [45]，以有监督的方式同时探索图像内和图像间的一致性。一些方法利用多任务学习来同时优化显著性检测和协同分割 [46] 或协同峰值搜索 [14]。其他工作从多尺度 (multi-scale) [47]、多阶段 (multi-stage) [48] 或多层 (multi-layer) [49] 特征探索更丰富的特征。另一个值得注意的研究方向是探索组级语义表征 (共识)，它用于检测每个图像的协同显著区域。获取辨别性语义表征的方法有多种，如群体注意力语义聚合 [50]、梯度反馈 [28]、同类别关联 (co-category association) [28]、联合完全卷积网络 (united fully convolutional network) [28] 或集成多层图 (integrated multilayer graph) [51]。一些方法以半监督 [52] 或无监督 [25, 53, 54, 55] 的方式解决 CoSOD，并对单幅图像的共显著性检测 [56, 57] 进行了研究。

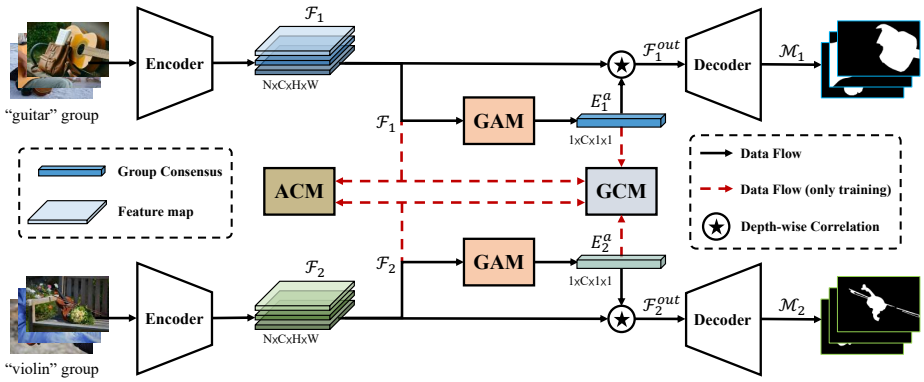

图 2. 组间协作学习网络 (GCoNet) 结构。两组图像首先由共享参数的编码器处理。然后，本文使用组内亲和模块 (GAM, 更多细节见图 3) 对每个组进行组内协作学习来生成共识，该共识与原始特征进行协作，其结果送入解码器来获得协同显著对象的分割结果。此外，两个组的原始特征图和共识被送到组间协作模块 (GCM，见图 4) 来进行组间协作学习。此外，还应用辅助分类模块 (ACM) 来获取高层语义表示。GCM 和 ACM 仅用于训练，在推理时会被移除。

以前的工作关注组内 (图像内和图像间) 线索，用于捕捉协同显著对象的协同属性。虽然 CODW [58] 利用了视觉上相似的物体图像，但组间信息没有受到太多关注。最近，Zhang 等人 [28] 利用拼图训练隐式地挖掘其他图像，以促进群组训练。但他们的模式仍然以组内学习为目标。本文的方法不同于现有的模型，它探索群组间的关系，显式地在图像组层面学习更好的特征。

3. 组间协作学习网络

3.1. 结构概述

给定一组包含某一类的协同显著目标对象的 N 张相关图片 $\{I_1, I_2, ..., I_n\}$ 组成图像组，CoSOD 同时检测并输出他们的显著图。不同于现有的仅依赖于图像组内信息的 CoSOD 方法，本文提出了一种新的组间协作学习网络 (GCoNet) 来挖掘组内和组间的共识特征。

图 2 是本文的 GCoNet 的流程图。首先，一个编码器网络用来提取两个图像组的特征图 $\mathcal{F}_1 = \{F_{1,n}\}_{n=1}^{N}, \mathcal{F}_2 = \{F_{2,n}\}_{n=1}^{N} \in \mathbb{R}^{N\times C\times H\times W}$，其中 C 是通道数量，$H \times W$ 是空间大小。然后，组内亲和模块 (GAM) 分别从图像组特征 $\mathcal{F}_1$, $\mathcal{F}_2$ 中提取得到共识特征 E_1^a 和 $E_2^a \in \mathbb{R}^{1\times C\times 1\times 1}$，它们表示每个图像组中协同显著对象的共有属性 (在本文的实验中 $C = 512$)。同时，应用组间协作模块 (GCM) 来增强图像表示，从而区分不同图像组之间的目标属性。最后，为了学习更好的特征嵌入空间，本文使用辅助分类模块 (ACM) 进一步提升了图像的高层语义表示。然后，将协同特征送到解码器网络来产生协同显著图 $\mathcal{M}_1, \mathcal{M}_2$。

3.2. 组内亲和模块

直观地说，同一类中的相同物体在外观上总是有一些相似之处，在特征上也有很高的相似性，这在许多任务中都得到了广泛的应用。受自监督视频跟踪方法 [59, 60, 61, 62] 的启发，该方法基于相邻帧之间的像素级别的对应关系来传播目标对象的分割掩模，本文通过计算组内图片的全局亲和力将这一思想扩展到 CoSOD 任务中。

对于任意两个图像特征 $\{F_{1,n}, F_{1,m}\} \in \mathcal{F}_1$ [2]，(为了简便，且不引入歧义的情况下，本文省去了表示图像组的下标)，使用内积来计算它们像素级别的相关关系：

$$S_{(n,m)} = \theta(F_n)^T \phi(F_m), \tag{1}$$

其中 θ, ϕ 是线性嵌入函数 ($3 \times 3 \times 512$ 的卷积层)。亲和注意力图 $S_{(n,m)} \in \mathbb{R}^{HW\times HW}$ 有效地捕捉了图像对 (n, m) 中协同显著物体的共性。然后，本文找到 F_n 上每个像素与 F_m 相关性的极大值，生成 F_n 的亲和力图 $A_{n\leftarrow m} \in \mathbb{R}^{HW\times 1}$，从而排除图中噪声相关值的影响。

同样地，本文可以把两个图像的局部亲和力扩展到该组中所有图像的全局亲和力。具体来说，本文用

[2] 在章节3.2中关于 $\mathcal{F}_1$ 的所有分析同样可以被用于 $\mathcal{F}_2$。为了简化符号，本文省略了下标。例如，本文用 F_n 来表示 $F_{1,n}$。

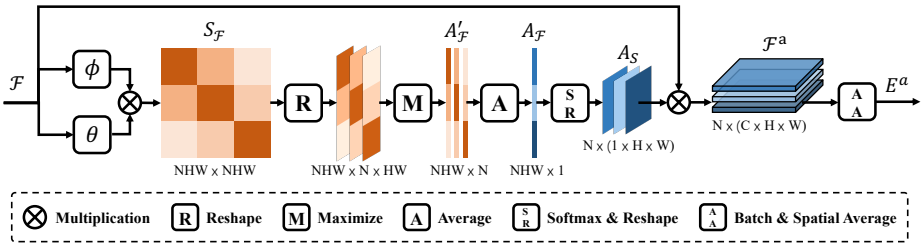

图 3. 组内亲和模块。本文首先利用亲和力注意，通过在组内协作所有图像来生成输入特征的注意力图。随后，将注意力图与输入特征相乘，以生成该组的共识。然后，将获得的共识用于协调原始特征图，并将其送入到 GCM 进行组间协作学习。

式 1. 计算所有图像特征 $\mathcal{F}$ 之间的亲和注意力图 $S_{\mathcal{F}} \in \mathbb{R}^{NHW \times NHW}$ 。然后，根据 $S_{\mathcal{F}}$ 求出每张图像的极大值 $A'_{\mathcal{F}} \in \mathbb{R}^{NHW \times N}$ ，并对 N 张图像的极大值求平均值来生成全局亲和注意力图 $A_{\mathcal{F}} \in \mathbb{R}^{NHW \times 1}$。这样，亲和注意力图在所有图像上进行全局优化，从而消除了偶然的共生偏差的影响。然后，使用 Softmax 操作对 $A_{\mathcal{F}}$ 进行归一化，并将其重塑成注意力图 $A_S \in \mathbb{R}^{N \times (1 \times H \times W)}$。本文将 A_S 与原始特征 $\mathcal{F}$ 相乘以产生注意力特征图 $\mathcal{F}^a \in \mathbb{R}^{N \times C \times H \times W}$。最后，对整个组的所有注意力特征图 $\mathcal{F}^a$ 沿 batch 和 spatial 维度进行平均池化来产生注意力共识 E^a，如图 3所示。

全局亲和力模块着重于捕捉同一组内协同显著对象之间的共性，因此提高了共识表示的组内紧凑性。这种组内紧凑性减轻了共同出现的噪声干扰，使模型能够关注协同显著对象所在的区域。这使得协同显著对象的共享属性能够被更好的捕捉，从而产生更好的共识表示。通过逐通道深度相关性操作 [63, 64]，将获得的注意力共识 E^a 与原始特征图 $\mathcal{F}$ 相结合以实现高效的信息关联。由此产生的特征图 $\mathcal{F}^{out}$ 被送入解码器，以预测每幅图像的协同显著图 $\mathcal{M}_n$。损失函数为:

$$\mathcal{L}_{\text{sal}} = \frac{1}{N} \sum_{n}^{N} \mathcal{L}_{\text{siou}}(\mathcal{M}_n, \mathcal{G}_n), \tag{2}$$

其中 $\mathcal{L}_{\text{siou}}$ 是 soft IoU 损失函数 [29, 65] 并且 $\mathcal{G}_n$ 表示图像组中每张图片的真实标签 (ground-truth)。

3.3. 组间协作模块 (GCM)

大多数 CoSOD 方法倾向于关注共识的组内紧凑性，但组间可分性对于区分干扰的物体同样至关重要，尤其是在处理有多个显著物体的复杂图像时。为了加强不同组之间的辨别表征，本文提出了一个简单而有效的模块，即 GCM，通过学习来编码组间的可分性。

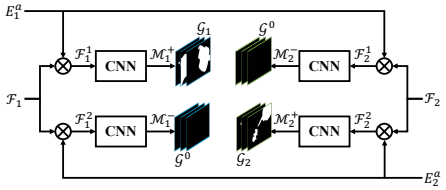

图 4. 组间协作模块。两组的原始特征图和共识被送入 GCM。GCM 中以一致的特征和共识 (来自同一图像组) 来预测的输出由真实标签 (ground-truth) 进行监督。否则，其输出将由全零的掩膜 (mask) 来监督。

给出两个图像组，其对应的特征分别为 $\{\mathcal{F}_1, \mathcal{F}_2\}$ 并且由 GAM 模块得到的注意力共识为 $\{E_1^a, E_2^a\}$，本文采用组内和组间交叉乘法对它们进行处理。具体来说，组内乘法处理的是特征和它们自身的共识: $\mathcal{F}_1^1 = \mathcal{F}_1 \cdot E_1^a$ 和 $\mathcal{F}_2^2 = \mathcal{F}_2 \cdot E_2^a$ 用于组内协作，而组间乘法则作用于不同组的特征和共识，$\mathcal{F}_1^2 = \mathcal{F}_1 \cdot E_2^a$ 和 $\mathcal{F}_2^1 = \mathcal{F}_2 \cdot E_1^a$，用于表示组间的交互。利用组内乘法特征 $\mathcal{F}^+ = \{\mathcal{F}_1^1, \mathcal{F}_2^2\}$ 来预测协同显著图，而组间乘法特征 $\mathcal{F}^- = \{\mathcal{F}_1^2, \mathcal{F}_2^1\}$ 用来提供具有组间可分性的共识监督。具体来说，本文将 $\{\mathcal{F}^+, \mathcal{F}^-\}$ 送入一个带有上采样层的小型卷积网络来产生显著图 $\{\mathcal{M}^+, \mathcal{M}^-\}$[3] 并以不同的监督信号监督 (本文使用真实标签来监督 $\mathcal{F}^+$，然而对于 $\mathcal{F}^-$ 本文使用的是全零标签监督。) 损失函数:

$$\mathcal{L}_{\text{ctm}} = \frac{1}{N} \sum_{n}^{N} \mathcal{L}_{\text{FL}}(< \mathcal{M}_n^+, \mathcal{M}_n^- >, < \mathcal{G}_n, \mathcal{G}_n^0 >), \tag{3}$$

其中 $\mathcal{L}_{\text{FL}}$ 是 focal loss [66], $\mathcal{G}_n$ 是 ground-truth, $\mathcal{G}_n^0$ 是全零掩膜，此外, $< \cdot >$ 代表并联操作。

因此，本文的 GCM 鼓励共识用较高的组间可分性来区分不同的组，以识别复杂环境中的干扰。另一个

[3] $\mathcal{M}^+ = \{\mathcal{M}_1^+, \mathcal{M}_2^+\}$ 和 $\mathcal{M}^- = \{\mathcal{M}_1^-, \mathcal{M}_2^-\}$。

表 1. 本文 GCoNet 中的 GAM (组间亲和模块)、GCM (组间协作模块)、ACM (辅助分类模块) 及其组合的有效性的**定量消融研究**。

ID	Modules			CoCA [28]				CoSOD3k [9]				Cosal2015 [58]			
	GAM	GCM	ACM	$E_\phi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$\epsilon\downarrow$	$E_\phi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$\epsilon\downarrow$	$E_\phi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$\epsilon\downarrow$
1				0.618	0.591	0.419	0.190	0.811	0.764	0.721	0.108	0.862	0.818	0.800	0.087
2	✓			0.663	0.605	0.442	0.160	0.823	0.772	0.736	0.099	0.873	0.825	0.815	0.079
3		✓		0.666	0.616	0.452	0.156	0.839	0.788	0.748	0.087	0.877	0.834	0.823	0.074
4			✓	0.651	0.606	0.442	0.167	0.829	0.779	0.737	0.094	0.875	0.832	0.820	0.076
5	✓	✓		0.719	0.650	0.504	0.126	0.850	0.798	0.766	0.078	0.884	0.842	0.837	0.070
	✓	✓	✓	**0.760**	**0.673**	**0.544**	**0.105**	**0.860**	**0.802**	**0.777**	**0.071**	**0.888**	**0.845**	**0.847**	**0.068**

优点是，这个模块使模型能够在现有的 SOD 数据集上进行训练，其图像通常只包含一个主导物体。本文可以在推理过程中丢弃这个模块从而不引入额外的计算量。

3.4. 辅助分类模块 (ACM)

为了获得更具鉴别性的共识特征，本文还引入了一个 ACM 来提升高层级的语义表征学习。具体来说，本文在骨干网络中加入一个带有全局平均池层和一个全连接层的分类预测器，将 F_n 分类到相应的类 $\mathcal{Y}_n$。在欧氏特征空间中，分类监督可以通过引入大的间距 (margin) 来分离类，并对属于同一类的样本进行聚类。因此，它能使模型产生更具代表性的特征，有利于共识学习到组内紧凑性和组间可分性。损失函数为:

$$\mathcal{L}_{\text{cls}} = \mathcal{L}_{\text{ce}}(\mathcal{Y}_n, \hat{\mathcal{Y}}_n), \tag{4}$$

其中 $\mathcal{L}_{\text{ce}}$ 是交叉熵损失，$\hat{\mathcal{Y}}_n$ 是真实类别。

3.5. 端到端训练

在训练过程中，GAM、GCM 和 ACM 以端到端的方式与骨干网联合训练。整个框架通过整合上述所有损失函数进行优化:

$$\mathcal{L} = \lambda_1\mathcal{L}_{\text{sal}} + \lambda_2\mathcal{L}_{\text{ctm}} + \lambda_3\mathcal{L}_{\text{cls}}, \tag{5}$$

其中 λ_1, λ_2, 和 λ_3 是用于平衡损失函数的超参数权重。

4. 实验

4.1. 实验细节

本文使用特征金字塔网络 (FPN) [67] 和 VGG-16 [68] 作为本文的主干。为了进行公平的比较，本文和 GICD [28] 一样使用 DUTS [69] 数据集作为本文的训练集。GICD [28] 中的分组标签用于在训练期间对图像进行分组。在每个训练集中，本文随机选择两个不同的组，每组 16 个[4]样本来训练网络。所有图像的大小都调整到 224×224，用于训练和测试，并且输出的显著图被还原到原图的大小以进行评估。该网络使用 Adam OptiMizer 进行了总共 50 个轮次的训练。初始学习速率设置为 $1e-4$，$\beta_1 = 0.9$ 和 $\beta_2 = 0.99$，并且学习率在每经过 25 个轮次除以 10。整体训练时间约为 4 个小时，每个图像对[5]的推理速度为 16 ms。训练和推理平台配备了 56 个 Intel (R) Xeon (R) CPU E5-2680 v4 @ 2.40GHz 和一个 Nvidia GeForce GTX 1080Ti。

[4]由于计算资源的限制，越大越好。

4.2. 评价数据集和评价指标

本文采用了三个具有挑战性的数据集进行评估: CoCA [28]、CoSOD3k [70] 和 Cosal2015 [58]。最后的是一个广泛应用于 CoSOD 方法评估的大数据集。前两个是最近提出的，用于真实世界的协同显著性方法评估，图像通常包含复杂背景下的多个共同和非共同对象。遵循最近的大规模基准工作的建议 [70]，本文没有使用 iCosig [71] 和 MSRC [72] 进行评估，因为其通常在一幅图像中只提供一个显著对象，不太适合评估 CoSOD 模型。在实验中，本文使用 maximum E-measure $E_\phi^{\max}$ [73]、S-measure S_α [74]、maximum F-measure $F_\beta^{\max}$ [75] 和 mean absolute error (MAE)ϵ [76] 来评估本文的模型[6]。

4.3. 消融实验

在本节中，本文将研究方法中每个模块的有效性 (表1)，并研究它们如何为形成良好的共识特征做贡献。**GAM 的有效性**。全局协同注意力模块是该模型的基本组成部分，旨在捕捉图像组中协同显著对象的共同属性，以获得更好的组内紧凑性。与通过平均池化操作仅提取普通共识 (vanilla consensus) 的基线 (baseline) 模型相比，GAM 提高了模型在所有评价指标和数据集

[5]CoSOD 任务对图像组起作用。因此，本文使用基本图像对组来评估速度，而不是单一图像。

[6]评估工具: https://github.com/DengPingFan/CoSODToolbox。

表 2. 本文的 GCoNet 和其他模型的**量化比较结果**。“↑” (“↓”) 代表越高 (低) 越好。Co = CoSOD 模型, Sin = Single-SOD 模型。符号 ∗ 代表传统的 CoSOD 方法。

Method	Pub. & Year	Type	CoCA [28]				CoSOD3k [9]				Cosal2015 [58]			
			$E_\phi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$\epsilon\downarrow$	$E_\phi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$\epsilon\downarrow$	$E_\phi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^{\max}\uparrow$	$\epsilon\downarrow$
CBCD* [21]	TIP 2013	Co	0.641	0.523	0.313	0.180	0.637	0.528	0.466	0.228	0.656	0.544	0.532	0.233
GWD [77]	IJCAI 2017	Co	0.701	0.602	0.408	0.166	0.777	0.716	0.649	0.147	0.802	0.744	0.706	0.148
RCAN [78]	IJCAI 2019	Co	0.702	0.616	0.422	0.160	0.808	0.744	0.688	0.130	0.842	0.779	0.764	0.126
CSMG [47]	CVPR 2019	Co	0.733	0.627	0.499	0.114	0.804	0.711	0.709	0.157	0.842	0.774	0.784	0.130
BASNet [29]	CVPR 2019	Sin	0.644	0.592	0.408	0.195	0.804	0.771	0.720	0.114	0.849	0.822	0.791	0.096
PoolNet [30]	CVPR 2019	Sin	0.640	0.602	0.404	0.177	0.799	0.771	0.709	0.113	0.848	0.823	0.785	0.094
EGNet [31]	ICCV 2019	Sin	0.648	0.603	0.404	0.178	0.793	0.762	0.702	0.119	0.843	0.818	0.786	0.099
SCRN [79]	ICCV 2019	Sin	0.642	0.612	0.413	0.164	0.805	0.771	0.716	0.113	0.850	0.817	0.783	0.098
GICD [28]	ECCV 2020	Co	0.715	0.658	0.513	0.126	0.848	0.797	0.770	0.079	**0.887**	0.844	0.844	0.071
CoEGNet [9]	TPAMI 2021	Co	0.717	0.612	0.493	0.106	0.825	0.762	0.736	0.092	0.882	0.836	0.832	0.077
GCoNet (本文)	CVPR 2021	Co	**0.760**	**0.673**	**0.544**	**0.105**	**0.860**	**0.802**	**0.777**	**0.071**	**0.887**	**0.845**	**0.847**	**0.068**

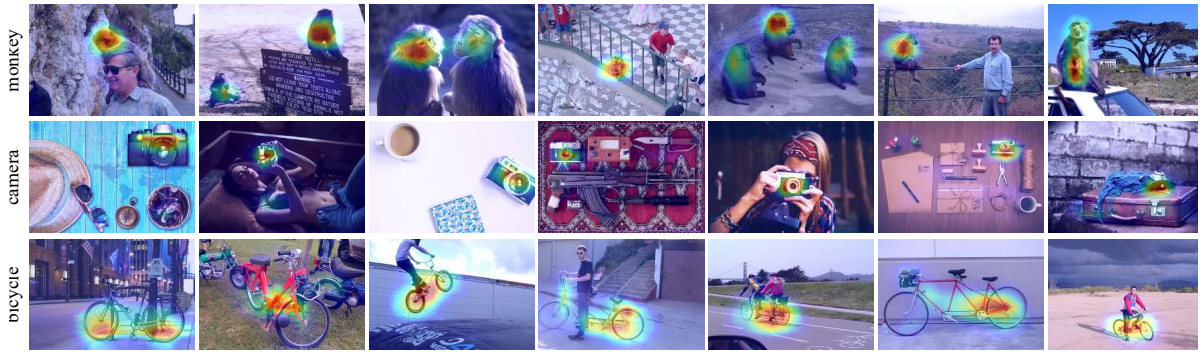

图 5. 使用组内协作学习在每组的所有图像中执行 GAM 学习到的**亲和力注意图的可视化**。生成的注意力掩膜对具有共享属性的协同显著区域很敏感，这有利于共识特征的学习。

的性能。为了更深入地理解本文的 GAM 模块，本文可视化了其学习到的注意力掩码, 如图 5所示。本文发现，全局协同注意力模块有效地缓解了图像组内的共生噪声, 并且聚焦在图像组中的协同显著区域，例如，在猴子和自行车组中，在一些图像中共同出现了人，但本文的 GAM 没有受到不利影响。GAM 能够从整个图像组的全局视角检测到全局协同对象，而局部配对协同注意不能在局部视图中区分它们。

GCM 的有效性。组间协作模块的设计是为了使共识具备组间可分性，从而能够区分组内非共同对象的干扰，在模型配置了 GCM 后，表1中的性能得到了显著的提高 (ID-1 versus ID-3)，特别是在具有挑战性的 CoCA [28] 数据集上，该数据集的图像通常包含多个非共同和共同的对象。为了研究使用 GCM 训练模型时的共识特征，本文在 CoCA 数据集上使用 t-SNE [19] 来可视化共识，并与没有使用 GCM 的普通共识进行了比较。如图 1所示，普通共识 (顶端: CoEGNet [9]) 倾向于聚集在一起，即使它们属于不同的组，导致协同显著性目标检测结果模糊，特别是对于那些相似但不同组的对象。相比之下，用 GCM 训练的共识 (底部: 本文的方法) 更加多样化，组间差异更大，可以更有效地实现组间分离。为了进行定量比较，本文评估了 ” 吉他 ” 和 ” 小提琴 ” 共识的余弦相似度 (↓ 越低越好)，本文的方法 (0.32) 比 CoEGNet (0.75) 好很多。

ACM 的有效性。如表 1所示, 分类模块使用辅助分类监督为共识特征提供了更好的骨干特征。ACM 提高了 baseline 在所有指标和数据集的性能。这种改进不会改变网络结构，也不会引入额外的计算开销，因此有很大的潜力来帮助其他模型和任务来利用多任务学习获得更具代表性的特征。

4.4. 与其他方法的比较

由于并不是所有的 CoSOD 模型都公开发布了代码，所以本文只将 GCoNet 与一个有代表性的传统算法 (CBCD) 和五个基于深度学习的 CoSOD 模型进行了比较，这些模型包括 GWD [80]、RCAN [78]、CSMG [47]、GICD [28] 和 CoEGNet [9]。根据目前最先进的模型 [28], 本文还与四种前沿的基于深度学习的

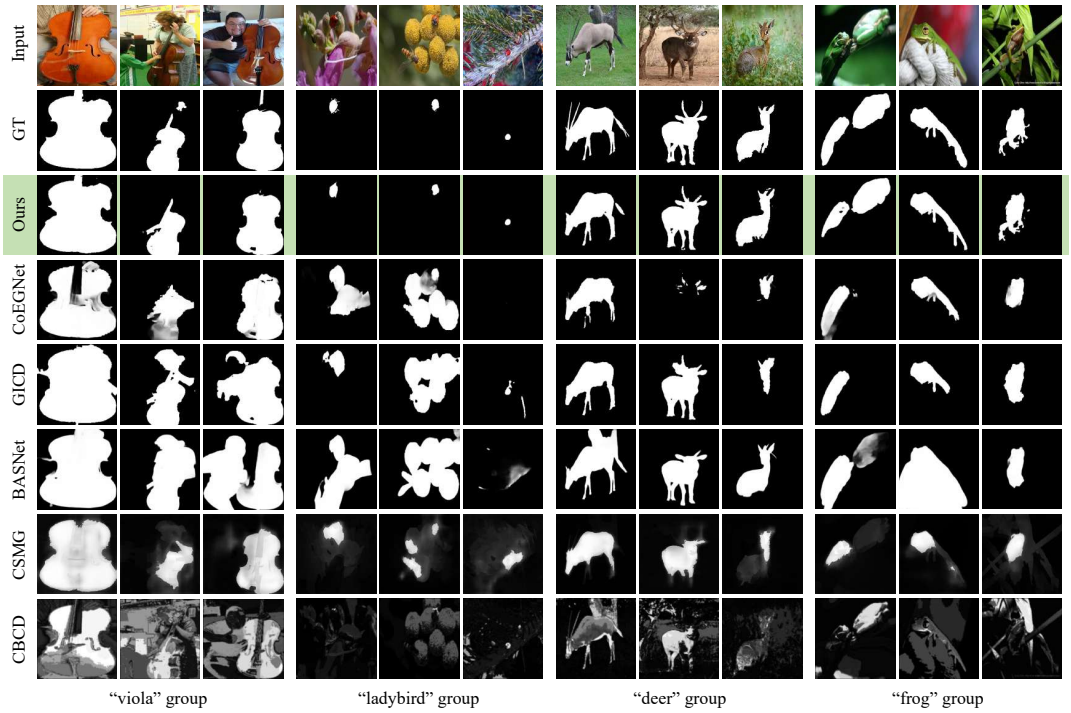

图 6. 本文的模型 GCoNet 和其他模型的**定性比较**。‘GT’ 代表 GroundTruth。

显著目标检测 (SOD)[7]模型进行了比较：BASNet [29]、PoolNet [30]、EGNet [31] 和 SCRN [79]。更完整的排行榜可以在最近的评测工作 [9, 70] 中找到。

量化结果。 表 2列出了本文的模型和最新方法的定量结果。本文的模型在所有指标上都优于它们，特别是在具有挑战性的 CoCA 和 CoSOD3k 数据集上。在这三个数据集中，CoCA 是最具挑战性的，因为除了尺寸更小的协同显著对象之外，图像通常还包含其他多个对象。本文的模型利用了更好的共识，并且明显优于其他方法，特别是 SOD 方法，这些方法受限于区分干扰对象。CoSOD3k 具有相似的特点，本文的模型在此数据集上的性能仍然比其他模型好得多。Cosal2015 是最简单的数据集，因为它的图像通常只包含一个协同显著的对象，因此 SOD 算法可以很容易地处理这个数据集。本文的模型在这个数据集上不能充分利用更好的共识特征，所以提升并不像在其他数据集上那么显著。

定性结果。 图 6 展示了由不同的方法定性比较生成的显著图。在这些困难样本中，除了协同显著对象之外，每个图像还包含其他多个对象。如上所述，SOD 方法只能检测到显著对象，由于其内在限制，无法区分协同显著的对象。CoSOD 方法比 SOD 方法表现更好，因为它们能够区分协同显著区域方面的共识特征。然而，受限于共识特征较弱，他们无法处理更具挑战性的情况。本文的模型通过优化组内紧凑性和组间可分性引入了有效的共识，因此在检测协同显著对象方面表现得更好。

[7]SOD 方法可以直接应用于 CoSOD 任务。

5. 关于模块协作的讨论

本文的三个模块紧密地相互依赖并相互加强，从而提高了协同显著目标检测的性能。与单独使用 GAM 和 GCM 模块相比，GAM 和 GCM 的结合使用可以显著提高模型性能。没有 GAM，普通共识就不能很好地抑制非协同目标和背景引起的噪声，并且低质量的共识不能充分利用 GCM (主要依赖于共识来区分不同对象) 来区分不同对象。另一方面，虽然共识可以在 GAM 的帮助下捕捉到共同的属性，但如果没有 GCM，很难区分相似的图像组。总体来说，GAM 产生的具有高组内紧凑性的较好共识来检测协同显著对象，而 GCM 则进一步赋予共识组间可分性，从而使其共识有更好的辨别性。在 ACM 中，共识可以受益于多任务学习，从而学习到更具代表性的特征。

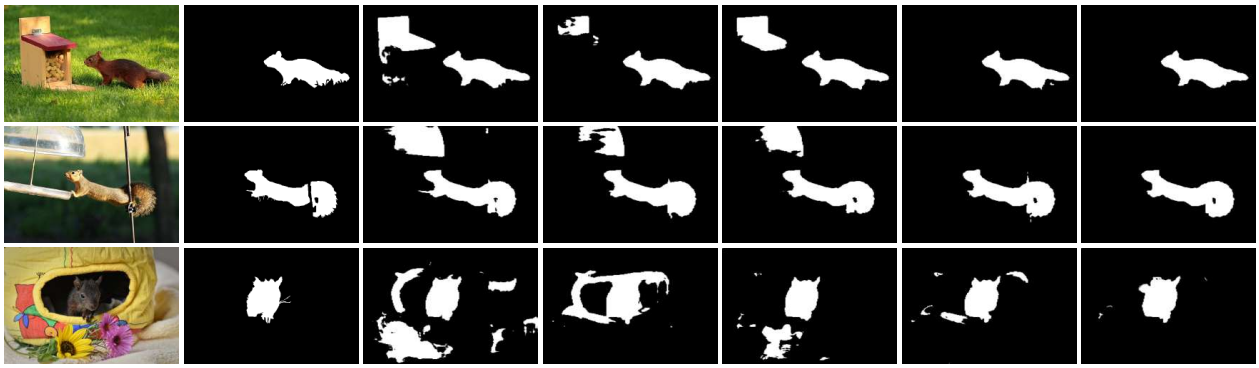

图 7. 本文的 GCoNet 在不同模块及其组合上的**定性消融实验**

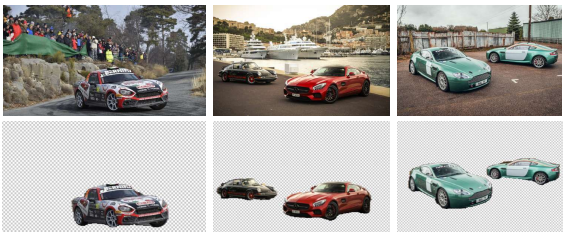

图 8. 应用 1。由本文的 GCoNet 获得的内容感知对象协同分割的可视化结果（“GT CAR”）。

图 9. 应用 2。本文的 GCoNet 生成的基于协同定位的自动缩略图（“浣熊”）。

图 7 定性分析它们的组合。Baseline 模型检测到了非协同目标，而 GAM 和 GCM 略有改善其不良影响，将 GAM 和 GCM 相结合，可以有效捕获共显著目标，ACM 进一步提高了共显著目标检测结果。

6. 下游应用

在这里，本文展示了如何利用提取的协同显著图来为选定的密切相关的下游图像处理任务生成高质量的分割掩码。

应用 #1: 内容感知 (Content-Aware) 协同分割。协同显著图已被用于无监督对象分割的预处理。在本文的实现中，首先通过关键字搜索从互联网上手动选择一组图像。然后，本文的 GCoNet 重新生成协同显著图来自动挖掘特定组的显著内容。与 Cheng [27] 等人相似，本文也用 GrabCut [81] 得到了最终的分割结果。要初始化 GrabCut，本文只需选择自适应阈值 [82] 来二值化显著图。图 8展示出了内容感知对象协同分割的结果，可用于需要背景替换的现有电子商务应用。

应用 #2: 自动缩略图。配对图像缩略图的概念源于 [12]。出于同样的目的 [8]，本文提出了一个基于 CNN 的照片分类应用程序，它对于在网站上与朋友分享照片很有价值。如图 9所示，本文首先根据 GCoNet 获得的协同显著图生成黄色方框。然后，简单地放大黄色方框，得到一个更大的红色方框。最后，本文采用了基于集合的图像裁剪技术 (collection-aware crops technique) [12] 来产生结果 (第二行)。

7. 结论

本文探讨了一种新的 CoSOD 组间协作学习框架 (GCoNet)。发现了组级共识可以引入有效的语义信息，有利于 CoSOD 任务中组内紧凑性和组间可分性的表征。本文的实验定量和定性地证明了本文的 GCoNet 的优点并且其性能优于现有的最先进的模型。此外，本文的 GCoNet 达到了实时的推理速度 (16ms)，这为许多应用提供了强大支撑，如协同分割、协同定位等。

[8] 请注意 Jacobs 等人的工作 [12] 仅限于图像对的情况。

References

[1] Qi Fan, Deng-Ping Fan, Huazhu Fu, Chi-Keung Tang, Ling Shao, and Yu-Wing Tai. Group collaborative learning for co-salient object detection. *arXiv: Computer Vision and Pattern Recognition*, 2021.

[2] Deng-Ping Fan, Ming-Ming Cheng, Jiang-Jiang Liu, Shang-Hua Gao, Qibin Hou, and Ali Borji. Salient objects in clutter: Bringing salient object detection to the foreground. In *ECCV*, pages 186–202, 2018.

[3] Mingchen Zhuge, Deng-Ping Fan, Nian Liu, Dingwen Zhang, Dong Xu, and Ling Shao. Salient object detection via integrity learning. *arXiv preprint arXiv:2101.07663*, 2021.

[4] Xuebin Qin, Deng-Ping Fan, Chenyang Huang, Cyril Diagne, Zichen Zhang, Adrià Cabeza Sant'Anna, Albert Suàrez, Martin Jagersand, and Ling Shao. Boundary-aware segmentation network for mobile and web applications. *arXiv preprint arXiv:2101.04704*, 2021.

[5] Tao Zhou, Deng-Ping Fan, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Rgb-d salient object detection: A survey. *CVM*, pages 1–33, 2021.

[6] Deng-Ping Fan, Zheng Lin, Zhao Zhang, Menglong Zhu, and Ming-Ming Cheng. Rethinking rgb-d salient object detection: Models, data sets, and large-scale benchmarks. *IEEE TNNLS*, 2021.

[7] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Saeed Anwar, Fatemeh Saleh, Sadegh Aliakbarian, and Nick Barnes. Uncertainty inspired rgb-d saliency detection. *arXiv preprint arXiv:2009.03075*, 2020.

[8] Zuyao Chen, Runmin Cong, Qianqian Xu, and Qingming Huang. Dpanet: Depth potentiality-aware gated attention network for rgb-d salient object detection. *TIP*, 2020.

[9] Deng-Ping Fan, Tengpeng Li, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Ming-Ming Cheng, Huazhu Fu, and Jianbing Shen. Re-thinking co-salient object detection. *TPAMI*, 2021.

[10] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu. Salientshape: group saliency in image collections. *TVC*, 30(4):443–453, 2014.

[11] Xiaochuan Wang, Xiaohui Liang, Bailin Yang, and Frederick WB Li. No-reference synthetic image quality assessment with convolutional neural network and local image saliency. *CVM*, 5(2):193–208, 2019.

[12] David E Jacobs, Dan B Goldman, and Eli Shechtman. Cosaliency: Where people look when comparing images. In *ACM UIST*, pages 219–228, 2010.

[13] Wenguan Wang and Jianbing Shen. Higher-order image co-segmentation. *TMM*, 18(6):1011–1021, 2016.

[14] Kuang-Jui Hsu, Yen-Yu Lin, and Yung-Yu Chuang. Deepco3: Deep instance co-segmentation by co-peak search and co-saliency detection. In *CVPR*, 2019.

[15] Yu Zeng, Yunzhi Zhuge, Huchuan Lu, and Lihe Zhang. Joint learning of saliency detection and weakly supervised semantic segmentation. In *ICCV*, 2019.

[16] Zhifan Gao, Chenchu Xu, Heye Zhang, Shuo Li, and Victor Hugo C de Albuquerque. Trustful internet of surveillance things based on deeply represented visual co-saliency detection. *IoT-J*, 7(5):4092–4100, 2020.

[17] Koteswar Rao Jerripothula, Jianfei Cai, and Junsong Yuan. Efficient video object co-localization with co-saliency activated tracklets. *TCSVT*, 29(3):744–755, 2018.

[18] Koteswar Rao Jerripothula, Jianfei Cai, and Junsong Yuan. Cats: Co-saliency activated tracklet selection for video co-localization. In *ECCV*, 2016.

[19] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 9(Nov):2579–2605, 2008.

[20] Hongliang Li and King Ngi Ngan. A co-saliency model of image pairs. *TIP*, 20(12):3365–3375, 2011.

[21] Huazhu Fu, Xiaochun Cao, and Zhuowen Tu. Cluster-based co-saliency detection. *TIP*, 22(10):3766–3778, 2013.

[22] Xiaochun Cao, Zhiqiang Tao, Bao Zhang, Huazhu Fu, and Wei Feng. Self-adaptively weighted co-saliency detection via rank constraint. *TIP*, 23(9):4175–4186, 2014.

[23] Dingwen Zhang, Deyu Meng, and Junwei Han. Co-saliency detection via a self-paced multiple-instance learning framework. *TPAMI*, 39(5):865–878, 2016.

[24] Junwei Han, Gong Cheng, Zhenpeng Li, and Dingwen Zhang. A unified metric learning-based framework for co-saliency detection. *TCSVT*, 28(10):2473–2483, 2017.

[25] Kuang-Jui Hsu, Chung-Chi Tsai, Yen-Yu Lin, Xiaoning Qian, and Yung-Yu Chuang. Unsupervised cnn-based co-saliency detection with graphical optimization. In *ECCV*, 2018.

[26] Wenbin Zou, Kidiyo Kpalma, Zhi Liu, and Joseph Ronsin. Segmentation driven low-rank matrix recovery for saliency detection. In *BMVC*, 2013.

[27] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu. Global contrast based salient region detection. *TPAMI*, 37(3):569–582, 2014.

[28] Zhao Zhang, Wenda Jin, Jun Xu, and Ming-Ming Cheng. Gradient-induced co-saliency detection. In *ECCV*, 2020.

[29] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *CVPR*, 2019.

[30] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang. A simple pooling-based design for real-time salient object detection. In *CVPR*, 2019.

[31] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet: Edge guidance network for salient object detection. In *ICCV*, 2019.

[32] Shang-Hua Gao, Yong-Qiang Tan, Ming-Ming Cheng, Chengze Lu, Yunpeng Chen, and Shuicheng Yan. Highly efficient salient object detection with 100k parameters. In *ECCV*, 2020.

[33] Kai Zhao, Shanghua Gao, Wenguan Wang, and Ming ming Cheng. Optimizing the F-measure for threshold-free salient object detection. In *ICCV*, 2019.

[34] Kai-Yueh Chang, Tyng-Luh Liu, and Shang-Hong Lai. From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model. In *CVPR*, 2011.

[35] Hwann-Tzong Chen. Preattentive co-saliency detection. In *ICIP*, 2010.

[36] Yijun Li, Keren Fu, Zhi Liu, and Jie Yang. Efficient saliency-model-guided visual co-saliency detection. *SPL*, 22(5):588–592, 2014.

[37] Zhi Liu, Wenbin Zou, Lina Li, Liquan Shen, and Olivier Le Meur. Co-saliency detection based on hierarchical segmentation. *SPL*, 21(1):88–92, 2013.

[38] Hongliang Li, Fanman Meng, and King Ngi Ngan. Co-salient object detection from multiple images. *TMM*, 15(8):1896–1909, 2013.

[39] Bo Jiang, Xingyue Jiang, Ajian Zhou, Jin Tang, and Bin Luo. A unified multiple graph learning and convolutional network model for co-saliency estimation. In *ACM MM*, 2019.

[40] Bo Jiang, Xingyue Jiang, Jin Tang, Bin Luo, and Shilei Huang. Multiple graph convolutional networks for co-saliency detection. In *ICME*, 2019.

[41] Kaihua Zhang, Tengpeng Li, Shiwen Shen, Bo Liu, Jin Chen, and Qingshan Liu. Adaptive graph convolutional network with attention graph clustering for co-saliency detection. In *CVPR*, 2020.

[42] Dingwen Zhang, Junwei Han, Jungong Han, and Ling Shao. Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining. *TNNLS*, 27(6):1163–1176, 2015.

[43] Wen-Da Jin, Jun Xu, Ming-Ming Cheng, Yi Zhang, and Wei Guo. Icnet: Intra-saliency correlation network for co-saliency detection. In *NeurIPS*, 2020.

[44] Koteswar Rao Jerripothula, Jianfei Cai, and Junsong Yuan. Quality-guided fusion-based co-saliency estimation for image co-segmentation and colocalization. *TMM*, 20(9):2466–2477, 2018.

[45] Xiwen Yao, Junwei Han, Dingwen Zhang, and Feiping Nie. Revisiting co-saliency detection: A novel approach based on two-stage multi-view spectral rotation co-clustering. *TIP*, 26(7):3196–3209, 2017.

[46] Chung-Chi Tsai, Weizhi Li, Kuang-Jui Hsu, Xiaoning Qian, and Yen-Yu Lin. Image co-saliency detection and co-segmentation via progressive joint optimization. *TIP*, 28(1):56–71, 2018.

[47] Kaihua Zhang, Tengpeng Li, Bo Liu, and Qingshan Liu. Co-saliency detection via mask-guided fully convolutional networks with multi-scale label smoothing. In *CVPR*, 2019.

[48] Min Li, Shizhong Dong, Kun Zhang, Zhifan Gao, Xi Wu, Heye Zhang, Guang Yang, and Shuo Li. Deep learning intra-image and inter-images features for co-saliency detection. In *BMVC*, 2018.

[49] Jingru Ren, Zhi Liu, Xiaofei Zhou, Cong Bai, and Guangling Sun. Co-saliency detection via integration of multi-layer convolutional features and inter-image propagation. *Neurocomputing*, 371:137–146, 2020.

[50] Qijian Zhang, Runmin Cong, Junhui Hou, Chongyi Li, and Yao Zhao. Coadnet: Collaborative aggregation-and-distribution networks for co-salient object detection. In *NeurIPS*, 2020.

[51] Dong-ju Jeong, Insung Hwang, and Nam Ik Cho. Co-salient object detection based on deep saliency networks and seed propagation over an integrated graph. *TIP*, 27(12):5866–5879, 2018.

[52] Xiaoju Zheng, Zheng-Jun Zha, and Liansheng Zhuang. A feature-adaptive semi-supervised framework for co-saliency detection. In *ACM MM*, 2018.

[53] Dingwen Zhang, Junwei Han, and Yu Zhang. Supervision by fusion: Towards unsupervised learning of deep salient object detector. In *ICCV*, 2017.

[54] Kuang-Jui Hsu, Yen-Yu Lin, and Yung-Yu Chuang. Co-attention cnns for unsupervised object co-segmentation. In *IJCAI*, 2018.

[55] Bo Li, Zhengxing Sun, Quan Wang, and Qian Li. Co-saliency detection based on hierarchical consistency. In *ACM MM*, 2019.

[56] Hongkai Yu, Kang Zheng, Jianwu Fang, Hao Guo, Wei Feng, and Song Wang. Co-saliency detection within a single image. In *AAAI*, 2018.

[57] Shaoyue Song, Hongkai Yu, Zhenjiang Miao, Dazhou Guo, Wei Ke, Cong Ma, and Song Wang. An easy-to-hard learning strategy for within-image co-saliency detection. *Neurocomputing*, 358:166–176, 2019.

[58] Dingwen Zhang, Junwei Han, Chao Li, Jingdong Wang, and Xuelong Li. Detection of co-salient objects by looking deep and wide. *IJCV*, 120(2):215–232, 2016.

[59] Carl Vondrick, Abhinav Shrivastava, Alireza Fathi, Sergio Guadarrama, and Kevin Murphy. Tracking emerges by colorizing videos. In *ECCV*, 2018.

[60] Zihang Lai and Weidi Xie. Self-supervised learning for video correspondence flow. In *BMVC*, 2019.

[61] Xiaolong Wang, Allan Jabri, and Alexei A Efros. Learning correspondence from the cycle-consistency of time. In *CVPR*, 2019.

[62] Zihang Lai, Erika Lu, and Weidi Xie. Mast: A memory-augmented self-supervised tracker. In *CVPR*, 2020.

[63] Bo Li, Wei Wu, Qiang Wang, Fangyi Zhang, Junliang Xing, and Junjie Yan. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In *CVPR*, 2019.

[64] Qi Fan, Wei Zhuo, Chi-Keung Tang, and Yu-Wing Tai. Few-shot object detection with attention-rpn and multi-relation detector. In *CVPR*, 2020.

[65] Zhuwen Li, Qifeng Chen, and Vladlen Koltun. Interactive image segmentation with latent diversity. In *CVPR*, 2018.

[66] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017.

[67] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017.

[68] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015.

[69] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *CVPR*, 2017.

[70] Deng-Ping Fan, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Huazhu Fu, and Ming-Ming Cheng. Taking a deeper look at the co-salient object detection. In *CVPR*, 2020.

[71] Dhruv Batra, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *CVPR*, 2010.

[72] John Winn, Antonio Criminisi, and Thomas Minka. Object categorization by learned universal visual dictionary. In *ICCV*, 2005.

[73] Deng-Ping Fan, Ge-Peng Ji, Xuebin Qin, and Ming-Ming Cheng. Cognitive vision inspired object segmentation metric and loss function. *SSI*, 2021.

[74] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A new way to evaluate foreground maps. In *ICCV*, 2017.

[75] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009.

[76] Ming-Ming Cheng, Jonathan Warrell, Wen-Yan Lin, Shuai Zheng, Vibhav Vineet, and Nigel Crook. Efficient salient region detection with soft image abstraction. In *ICCV*, 2013.

[77] Lina Wei, Shanshan Zhao, Omar El Farouk Bourahla, Xi Li, and Fei Wu. Group-wise deep co-saliency detection. In *IJCAI*, 2017.

[78] Bo Li, Zhengxing Sun, Lv Tang, Yunhan Sun, and Jinlong Shi. Detecting robust co-saliency with recurrent co-attention neural network. In *IJCAI*, 2019.

[79] Zhe Wu, Li Su, and Qingming Huang. Stacked cross refinement network for edge-aware salient object detection. In *ICCV*, 2019.

[80] Bo Li, Zhengxing Sun, Qian Li, Yunjie Wu, and Anqi Hu. Group-wise deep object co-segmentation with co-attention recurrent neural network. In *ICCV*, 2019.

[81] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Interactive foreground extraction using iterated graph cuts. *ACM TOG*, 23:3, 2012.

[82] Houwen Peng, Bing Li, Weihua Xiong, Weiming Hu, and Rongrong Ji. Rgbd salient object detection: a benchmark and algorithms. In *ECCV*, 2014.