

# 鐵路不是火車：顯著性作為弱監督語義分割的偽像素監督

Seungho Lee\*

延世大學

seungholee@yonsei.ac.kr

Minhyun Lee\*

延世大學

lmh315@yonsei.ac.kr

Jongwuk Lee

成均館大學

jongwuklee@skku.edu

Hyunjung Shim†

延世大學

kateshim@yonsei.ac.kr

## Abstract

現有使用圖像級弱監督的弱監督語義分割 (WSSS) 研究存在幾個限制：稀疏的物體覆蓋、不準確的物體邊界以及來自非目標物體的共現像素。為了克服這些挑戰，我們提出了一個新穎的框架，即顯式偽像素監督 (EPS)，通過結合兩種弱監督從像素級反饋中學習；圖像級標籤通過定位圖提供物體身份，來自現成顯著性檢測模型的顯著性圖提供豐富的邊界。我們設計了一種聯合訓練策略，以充分利用兩種信息之間的互補關係。我們的方法可以獲得準確的物體邊界並丟棄共現像素，從而顯著提高偽掩碼的質量。實驗結果表明，所提出的方法通過解決 WSSS 的關鍵挑戰顯著優於現有方法，並在 PASCAL VOC 2012 和 MS COCO 2014 數據集上實現了新的最先進性能。代碼可在 <https://github.com/halbielee/EPS> 獲得。

## 1. 介紹

弱監督語義分割 (WSSS) 利用弱監督（例如，圖像級標籤 [36, 37]、塗鴉 [29] 或邊界框 [22]）並旨在實現與完全監督模型相媲美的性能，這需要像素級標籤。大多數現有研究採用圖像級標籤作為分割模型的弱監督。WSSS 的整體流程由兩個階段組成。首先，使用圖像分類器為目標物體生成偽掩碼。然後，使用偽掩碼作為監督訓練分割模型。生成偽掩碼的流行技術是類激活映射 (CAM) [52]，它提供

\*表示相同貢獻。

†Hyunjung Shim 是通訊作者。

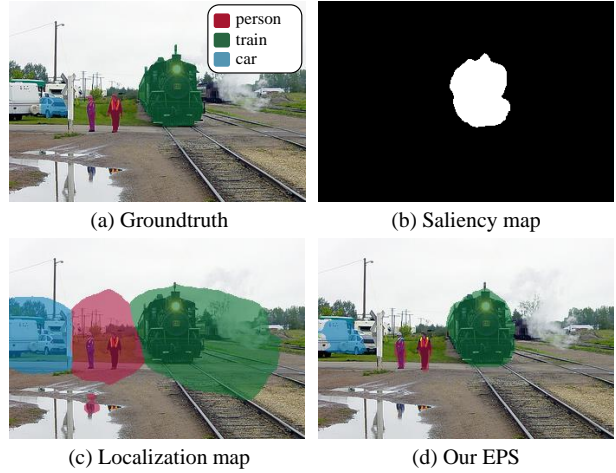


图 1. 利用顯著性圖和定位圖進行 WSSS 的激勵範例。(a) 真實值, (b) 通過 PFAN [51] 的顯著性圖, (c) 通過 CAM [52] 的定位圖, 以及 (d) 我們的 EPS 利用顯著性圖和定位圖來訓練分類器。請注意, 顯著性圖無法捕捉到人和車, 而我們的結果可以正確恢復它們, 並且定位圖過度捕捉了兩個物體。

與其圖像級標籤對應的物體定位圖。由於完全（即像素級註釋）和弱（即圖像級標籤）監督語義分割之間的監督差距，WSSS 存在以下關鍵挑戰：1) 定位圖僅捕獲目標物體的一小部分 [52]，2) 它遭受物體邊界不匹配的困擾 [23]，以及 3) 它幾乎無法將共現像素與目標物體分開（例如，鐵路與火車）[25]。

為了解決這些問題，現有的研究可以分為三個支柱。第一種方法是通過擦除像素 [9, 23, 28]、集成得分圖 [21, 27] 或使用自監督信號 [41] 來擴展物體覆蓋範圍以捕捉物體的全部範圍。然而，由於缺乏引導物體形狀的線索，它們無法確定目標物體的準確邊界。第二種方法專注於改善偽掩碼的物體邊界 [13, 32]。由於它們有效地學習了物體邊界，因此

自然地將偽掩膜擴展到邊界。然而，它們仍然無法區分非目標物體與目標物體的重合像素。這是因為前景和背景之間的強相關性（即，共現）幾乎無法從歸納偏差中區分出來（即，觀察到目標物體及其重合像素的頻率），如 [10] 所示。最後，第三種方法旨在使用額外的真實掩膜 [24] 或顯著性圖 [35, 47] 來緩解共現問題。然而，[24, 28] 需要強像素級註釋，這與弱監督學習範式相去甚遠。[35] 對顯著性圖的錯誤非常敏感。此外，[47] 無法覆蓋物體的全部範圍，並且存在邊界不匹配的問題。

在本文中，我們的目標是通過充分利用定位圖（即，使用圖像級標籤訓練的圖像分類器的 CAM）和顯著性圖（即，現成的顯著性檢測模型的輸出 [18, 34, 51]）來克服 WSSS 的三個挑戰。我們專注於定位圖和顯著性圖之間的互補關係。如圖 1 所示，定位圖可以區分不同的物體，但無法有效地分離它們的邊界。相反，雖然顯著性圖提供了豐富的邊界信息，但它並未揭示物體的身份。在這個意義上，我們認為我們的方法使用兩個互補的信息可以解決 WSSS 的性能瓶頸。

為此，我們提出了一種新的 WSSS 框架，稱為顯式偽像素監督（EPS）。為了充分利用顯著性圖（即，前景和背景），我們設計了一個分類器來預測  $C + 1$  類，包含  $C$  個目標類和背景類。我們利用  $C$  個定位圖和背景定位圖來估計顯著性圖。然後，顯著性損失被定義為顯著性圖和我們估計的顯著性圖之間的像素差異。通過引入顯著性損失，模型可以通過所有類的偽像素反饋進行監督。我們還使用多標籤分類損失來預測圖像級標籤。因此，我們訓練分類器以優化顯著性損失和多標籤分類損失，協同優化背景和前景像素的預測——我們發現我們的策略可以改善顯著性圖（第 3.3 節和圖 3）和偽掩膜（第 5.1 節和圖 4）。

我們強調，由於顯著性損失通過偽像素反饋懲罰邊界不匹配，它可以強制我們的方法學習物體的準確邊界。作為副產品，我們還可以通過將地圖擴展到邊界來捕捉整個物體。由於顯著性損失有助於將前景（例如，火車）與背景分開，我們的方法可以將共現像素（例如，鐵路）分配到背景類。實驗結果

表明，我們的 EPS 在 PASCAL VOC 2012 和 MS COCO 2014 數據集上取得了顯著的分割性能，創造了新的最先進的準確性記錄。

## 2. 相關工作

**弱監督語義分割。** WSSS 的一般流程是從分類網絡生成偽掩碼，並使用偽掩碼作為監督來訓練分割網絡。由於圖像級標籤中邊界信息的稀缺，許多現有方法遭受於不準確的偽掩碼。為了解決這個問題，交叉圖像親和性 [15]、知識圖譜 [31] 和對比優化 [38, 50] 被用來提高偽掩碼的質量。[5] 提出了一個自監督任務來發現子類別，以強化分類器來改進 CAM。[1, 2] 通過計算像素之間的親和性來隱式利用邊界信息。[49] 專注於生成可靠的像素級註釋，並設計了一個端到端的網絡來生成分割圖。[20, 25] 通過利用邊界損失來訓練分割網絡。最近，[3] 使用了一個基於單一分割的模型，並採用自監督訓練方案。[14] 通過利用多個不完整的偽掩碼來專注於分割網絡的魯棒性。

**基於顯著性的語義分割。** 顯著性檢測（SD）方法通過具有像素級註釋的外部顯著性數據集 [18, 46, 51] 或圖像級註釋 [39] 生成區分圖像中前景和背景的顯著性圖。許多 WSSS 方法 [15, 20, 27, 28, 42, 44] 利用顯著性圖作為偽掩碼的背景線索。[43] 利用顯著性圖作為單物體圖像的完全監督。[16] 使用實例級顯著性圖來學習物體的相似性圖。[6, 40, 47] 將顯著性圖與類別特定的注意力線索結合起來生成可靠的偽掩碼。[48] 使用單一網絡聯合解決 WSSS 和 SD，以提高兩個任務的性能。我們的 EPS 可以歸類為基於顯著性的方法，但在以下原因中明顯區別於其他方法。大多數現有方法將顯著性圖作為偽掩碼的一部分或作為改進分類器中間特徵的隱式指導。相反，我們的方法利用顯著性圖作為定位圖的偽像素反饋。儘管 [48] 在利用兩種互補信息的意義上與我們的工作最為相似，但他們既沒有解決共現問題，也沒有處理噪聲顯著性圖問題。

## 3. 提出的方法

在本節中，我們提出了一個新的弱監督語義分割（WSSS）框架，稱為顯式偽像素監督（EPS）。考

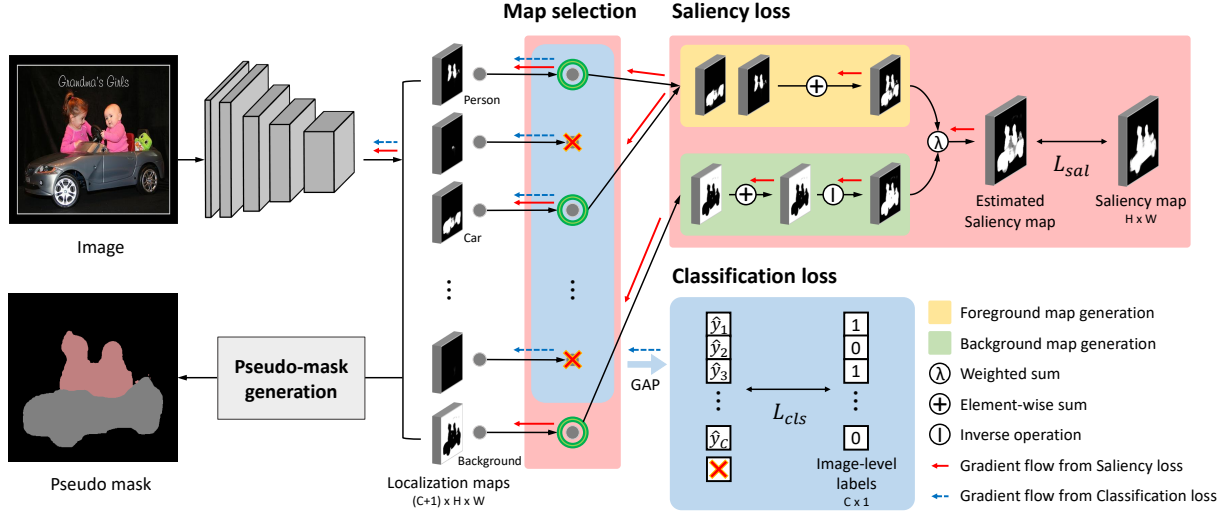


图 2. 我們的 EPS 整體框架。從骨幹網絡生成  $C + 1$  個定位圖。實際的顯著性圖是從現成的顯著性檢測模型生成的。某些目標標籤的定位圖被選擇性地用於生成估計的顯著性圖（第 3.2 節）。整體框架與顯著性損失和分類損失共同訓練（第 3.3 節）。

慮到 WSSS 的兩個階段，第一階段是生成偽掩碼，第二階段是訓練分割模型。在這裡，我們的主要貢獻是生成準確的偽掩碼。遵循 WSSS 的慣例 [13, 21, 27, 28, 41, 42]，我們然後訓練一個分割模型，其中第一階段生成的偽掩碼用作監督。

### 3.1. 動機

EPS 的關鍵見解是充分利用兩種信息的互補性，即來自定位圖的物體身份和來自顯著性圖的邊界信息。為此，我們利用顯著性圖作為定位圖的偽像素反饋，針對目標標籤和背景。我們設計了一個具有額外背景類別的分類器，從而預測總共  $C + 1$  類別，如圖 2 所示。使用該分類器，我們可以學習  $C + 1$  個定位圖，即  $C$  個目標標籤的定位圖和一個背景定位圖。我們接著解釋 EPS 如何解決 WSSS 中的邊界不匹配和共現問題。為了解決邊界不匹配問題，我們從  $C$  定位圖估計前景圖，並將其與顯著性圖的前景匹配。這樣，目標標籤的定位圖可以從顯著性圖中接收偽像素反饋，從而改善物體的邊界。為了減少非目標物體的共現像素，我們還將背景的定位圖與顯著性圖匹配。由於背景的定位圖也從顯著性圖中接收偽像素反饋，共現像素可以成功地分配給背景；非目標物體的共現像素大多與背景重疊。這就是為什麼我們的方法可以將共現像素從目標物體中分離出來。最後，EPS 的目標函數由兩部分組成：

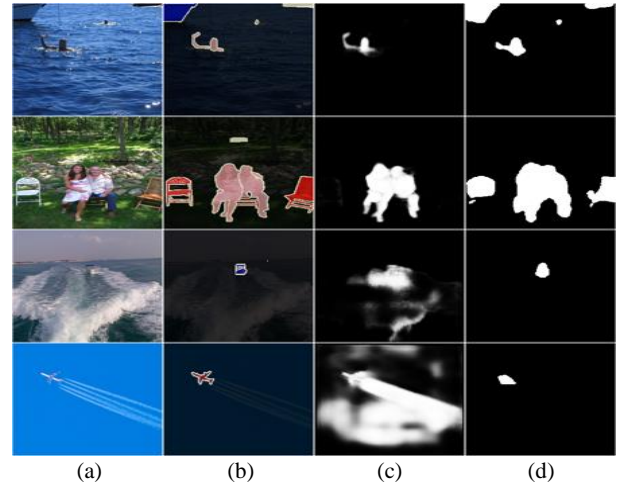


图 3. 在 PASCAL VOC 2012 上估計的顯著性圖的質性範例。(a) 輸入圖像，(b) 真實值，(c) 來自 [51] 的顯著性圖和 (d) 我們估計的顯著性圖。

通過顯著性圖的顯著性損失  $\mathcal{L}_{sal}$ （在圖 2 中以紅色框/箭頭標示）和通過圖像級標籤的多標籤分類損失  $\mathcal{L}_{cls}$ （在圖 2 中以藍色框/箭頭標示）。通過聯合訓練這兩個目標，我們可以將定位圖和顯著性圖的互補信息協同起來——我們觀察到通過我們的聯合訓練策略，彼此的噪聲和缺失信息得到了補充，如圖 3 所示。例如，從現成模型 [18, 34, 51] 獲得的原始顯著性圖存在缺失和噪聲信息。另一方面，我們的結果成功恢復了缺失的物體（例如，船或椅子）並去除了噪聲（例如，水泡或飛機尾跡），這顯然比原始顯



著性圖更好。因此，EPS 可以捕捉到更準確的物體邊界並將共現像素從目標物體中分離出來。這些優勢帶來了顯著的性能提升；表 6 報告稱，EPS 在分割準確性方面顯著超過現有模型，達到 3.8–10.6

### 3.2. 顯式偽像素監督

我們解釋如何利用顯著性圖進行偽像素監督。顯著性圖的關鍵優勢在於提供物體輪廓，這可以更好地揭示物體邊界。為了利用這一特性，我們將顯著性圖與兩種情況匹配：前景和背景。為了使類別定位圖與顯著性圖可比，我們合併目標標籤的定位圖並生成前景圖， $M_{fg} \in \mathbb{R}^{H \times W}$ 。我們也可以通過對背景圖進行反轉來表示前景，背景圖是背景標籤的定位圖  $M_{bg} \in \mathbb{R}^{H \times W}$ 。（稍後，我們將解釋如何細化前景圖以解決噪聲顯著性圖問題。）

具體來說，我們使用  $M_{fg}$  和  $M_{bg}$  估計顯著性圖  $\hat{M}_s$  如下：

$$\hat{M}_s = \lambda M_{fg} + (1 - \lambda)(1 - M_{bg}), \quad (1)$$

其中  $\lambda \in [0, 1]$  是調整前景圖和背景圖反轉加權和的超參數。（在我們的實驗中，默認將  $\lambda$  設置為 0.5，並在補充材料中提供了  $\lambda$  的額外消融研究。）然後，我們將顯著性損失  $\mathcal{L}_{sal}$  定義為我們估計的顯著性圖和實際顯著性圖之間的像素差異之和。（ $\mathcal{L}_{sal}$  的正式定義在第 3.3 節中給出。）

值得注意的是，使用預訓練模型被視為弱監督學習，因此利用顯著性圖已被廣泛接受為 WSSS 中的常見做法。儘管如此，採用完全監督的顯著性檢測模型可能存在爭議，因為它們使用來自不同數據集的像素級註釋。在本文中，我們研究了不同顯著性檢測方法的效果；1) 無監督和 2) 完全監督的顯著性檢測模型（見第 5.3 節），並實證表明我們的方法使用其中任何一種都優於所有其他方法 [13, 21, 40, 43, 47] 使用完全監督的顯著性模型。儘管現有方法在充分利用顯著性圖方面有限，我們的方法將顯著性圖作為偽像素監督，並將其作為邊界和共現像素的線索加以利用。

**處理顯著性偏差的地圖選擇。**之前，我們假設前景地圖可以是目標標籤的定位地圖的聯集；背景地圖可以是背景標籤的定位地圖。然而，這樣的簡單選擇規

則可能與現成模型計算的顯著性地圖不兼容。例如，來自 [51] 的顯著性地圖經常忽略一些物體作為顯著物體（例如，圖 1 中火車旁的小人物）。這種系統性錯誤是不可避免的，因為顯著性模型學習了不同數據集的統計數據。除非考慮到這個錯誤，否則相同的錯誤可能會傳播到我們的模型中，導致性能下降。

為了解決系統性錯誤，我們開發了一種有效的策略，使用定位地圖和顯著性地圖之間的重疊率。具體來說，如果第  $i$  個定位地圖  $M_i$  與顯著性地圖的重疊率超過  $\tau\%$ ，則將其分配給前景，否則分配給背景。形式上，前景和背景地圖計算如下：

$$\begin{aligned} M_{fg} &= \sum_{i=1}^C y_i \cdot M_i \cdot \mathbb{1}[\mathcal{O}(M_i, M_s) > \tau], \\ M_{bg} &= \sum_{i=1}^C y_i \cdot M_i \cdot \mathbb{1}[\mathcal{O}(M_i, M_s) \leq \tau] + M_{C+1}, \end{aligned} \quad (2)$$

其中  $y \in \mathbb{R}^C$  是二進制圖像級標籤， $\mathcal{O}(M_i, M_s)$  是計算  $M_i$  和  $M_s$  之間重疊率的函數。為此，我們首先將定位地圖和顯著性地圖二值化：對於像素  $p$ ， $B_k(p) = 1$  如果  $M_k(p) > 0.5$ ；否則  $B_k(p) = 0$ 。 $B_i$  和  $B_s$  分別是對應於  $M_i$  和  $M_s$  的二值化地圖。然後，我們計算  $M_i$  和  $M_s$  之間的重疊率，即  $\mathcal{O}(M_i, M_s) = |B_i \cap B_s| / |B_i|$ 。我們設置  $\tau = 0.4$ ，無論數據集和骨幹模型如何。在補充材料中，我們展示了我們的方法對  $\tau$  的選擇具有魯棒性（即， $\tau$  在  $[0.3, 0.5]$  之間顯示出可比的性能）。

我們將背景標籤的單一定位地圖與未選為前景的定位地圖結合起來。雖然這很簡單，但我們可以繞過顯著性地圖的錯誤，有效地訓練顯著性地圖忽略的一些物體。（在表 3 中，我們報告了所提出策略克服顯著性地圖錯誤的有效性。）

### 3.3. 聯合訓練過程

使用顯著性地圖和圖像級標籤，EPS 的整體訓練目標由兩部分組成，顯著性損失  $\mathcal{L}_{sal}$  和分類損失  $\mathcal{L}_{cls}$ 。首先，顯著性損失  $\mathcal{L}_{sal}$  通過測量實際顯著性地圖  $M_s$  和估計顯著性地圖  $\hat{M}_s$  之間的平均像素級距離來制定。

$$\mathcal{L}_{sal} = \frac{1}{H \cdot W} \|M_s - \hat{M}_s\|^2, \quad (3)$$

其中  $M_s$  是從現成的顯著性檢測模型——在 DUTS 數據集 [39] 上訓練的 PFAN [51] 獲得的。請注意，我們的方法無論顯著性檢測模型如何，一貫優於所有先前的藝術。

接下來，分類損失是通過圖像級標籤  $y$  和其預測  $\hat{y} \in \mathbb{R}^C$  之間的多標籤軟邊緣損失計算的，這是每個目標類的定位地圖上的全局平均池化的結果。

$$\mathcal{L}_{cls} = -\frac{1}{C} \sum_{i=1}^C y_i \log \sigma(\hat{y}_i) + (1 - y_i) \log (1 - \sigma(\hat{y}_i)), \quad (4)$$

其中  $\sigma(\cdot)$  是 sigmoid 函數。最後，總訓練損失是多標籤分類損失和顯著性損失的總和，即  $\mathcal{L}_{total} = \mathcal{L}_{cls} + \mathcal{L}_{sal}$ 。如圖 2 所示， $\mathcal{L}_{sal}$  涉及更新  $C + 1$  類別的參數，包括目標物體和背景。同時， $\mathcal{L}_{cls}$  僅評估  $C$  類別的標籤預測，不包括背景類別——來自  $\mathcal{L}_{cls}$  的梯度不會流入背景類別。然而，背景類別的預測可以被  $\mathcal{L}_{cls}$  隱式影響，因為它監督分類器的訓練。

#### 4. 實驗設置

**數據集。**我們在兩個流行的基準數據集上進行實證研究，分別是 PASCAL VOC 2012 [12] 和 MS COCO 2014 [30]。PASCAL VOC 2012 包含 21 個類別（即 20 個物體和背景），訓練集、驗證集和測試集分別有 1,464、1,449 和 1,456 張圖像。按照語義分割的常見做法，我們使用擴充的訓練集，其中包含 10,582 張圖像 [17]。接下來，COCO 2014 包含 81 個類別，包括一個背景，訓練和驗證集分別有 82,081 和 40,137 張圖像，其中不包含目標類別的圖像已被排除，如 [9] 所述。由於某些物體的真实分割標籤相互重疊，我們採用 COCO-Stuff [4] 的真实分割標籤，該標籤解決了同一 COCO 數據集上的重疊問題。

**評估協議。**我們在 PASCAL VOC 2012 的驗證集和測試集以及 COCO 2014 的驗證集上驗證我們的方法。PASCAL VOC 2012 測試集的評估結果來自官方 PASCAL VOC 評估服務器。此外，我們採用平均交集-聯合 (mIoU) 來衡量分割模型的準確性。

**實現細節。**我們選擇 ResNet38 [45] 作為我們方法的骨幹網絡，輸出步幅為 8。所有骨幹模型均在 ImageNet [11] 上進行預訓練。我們使用批量大小為 8 的 SGD 優化器。我們的方法訓練到 20k 次迭代，

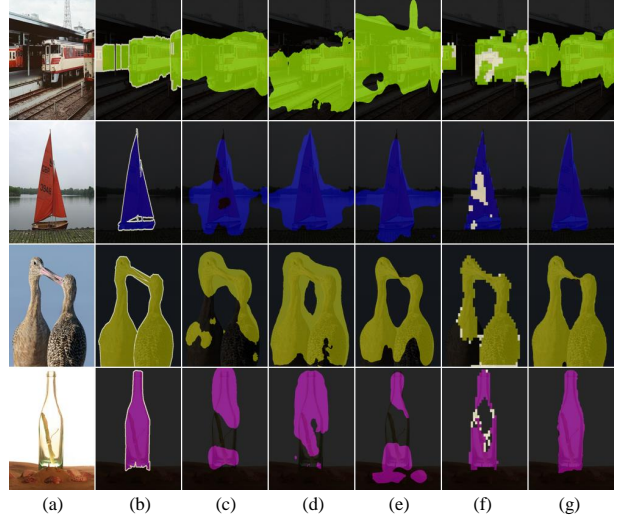


圖 4. PASCAL VOC 2012 上偽遮罩的質量比較。(a) 輸入圖像, (b) 真實值, (c) CAM, (d) SEAM, (e) ICD, (f) SGAN 和 (g) 我們的 EPS。

學習率為 0.01（最後一個卷積層為 0.1）。對於數據增強，我們使用隨機縮放、隨機翻轉和隨機裁剪到  $448 \times 448$ 。對於分割網絡，我們採用 DeepLab-LargeFOV (V1) [7] 和 DeepLab-ASPP (V2) [8]，以及 VGG16 和 ResNet101 作為其骨幹網絡。具體來說，我們使用四個分割網絡：基於 VGG16 的 DeepLab-V1 和 DeepLab-V2，基於 ResNet101 的 DeepLab-V1 和 DeepLab-V2。更詳細的設置在補充材料中。

#### 5. 實驗結果

##### 5.1. 處理邊界和共現問題

**邊界不匹配問題。**為了驗證偽掩碼的邊界，我們將邊界的質量與最先進的方法 [32, 41, 52] 進行比較。我們利用 SBD [17]，該數據集提供了邊界註釋和 PASCAL VOC 2011 的邊界基準。如 [32] 所述，邊界的質量以類別無關的方式進行評估，通過計算偽掩碼的邊緣來自拉普拉斯邊緣檢測器。然後，通過測量召回率、精確率和 F1 分數來評估邊界質量，將預測的邊界與真實邊界進行比較。表 1 報告了我們的方法在所有三個指標上大大超過其他方法。圖 4 中的定性示例顯示，我們的方法可以比所有其他方法捕捉到更準確的邊界。

**共現問題。**如幾項研究中所討論的 [20, 25, 28, 35]，我

方法	召回率 (%)	精確率 (%)	F1-分數 (%)	Baseline	Na"ive	Pre-defined	Our adaptive
CAM [52] <sub>CVPR'16</sub>	22.3	35.8	27.5	mIoU	66.1	66.5	67.9
SEAM [41] <sub>CVPR'20</sub>	40.2	45.0	42.5				
BES [32] <sub>ECCV'20</sub>	45.5	46.4	45.9				
我們的 EPS	60.0	73.1	65.9				

表 1. 在 SBD trainval 集上評估的邊界準確性。請注意, BES 的結果是從 [32]中提出的邊界預測網絡測量的。

方法	船 w/ 水	火車 w/ 鐵路	火車 w/ 平台
CAM [52] <sub>CVPR'16</sub>	0.74 (33.1)	0.11 (52.9)	0.09 (49.6)
SEAM [41] <sub>CVPR'20</sub>	1.13 (30.7)	0.24 (48.6)	0.20 (45.5)
ICD [13] <sub>CVPR'20</sub>	0.47 (41.4)	0.11 (56.7)	0.09 (49.2)
SGAN [47] <sub>ACCESS'20</sub>	0.10 (42.3)	0.02 (48.8)	0.01 (36.3)
我們的 EPS	0.10 (55.0)	0.02 (78.1)	0.01 (73.0)

表 2. 與現有代表性方法處理共現問題的比較。每個條目是 $m_{k,c}$  在 藍色 中 (越低越好), 括號中的 IoU (越高越好)。

們觀察到在 PASCAL VOC 2012 中, 一些背景類別經常與目標物體一起出現。我們通過使用 PASCAL-CONTEXT 數據集 [33] 進行定量分析, 該數據集提供了整個場景的像素級註釋 (例如, 水和 鐵路)。我們選擇了三對共現對; 船與 水, 火車與 鐵路, 以及火車與 平台。我們比較目標類別的 IoU 和目標類別與其共現類別之間的 混淆比率。混淆比率衡量共現類別被錯誤預測為目標類別的程度。混淆比率  $m_{k,c}$  通過  $m_{k,c} = FP_{k,c}/TP_c$  計算, 其中  $FP_{k,c}$  是共現類別  $k$  被錯誤分類為目標類別  $c$  的像素數,  $TP_c$  是目標類別  $c$  的真陽性像素數。關於共現問題的更詳細分析在補充材料中。Table 2 報告指出, EPS 一直顯示出比其他方法更低的混淆比率。SGAN [47] 的混淆比率與我們的方法相當相似, 但我們的方法在 IoU 方面更準確地捕捉到目標類別。有趣的是, SEAM 顯示出高混淆比率, 甚至比 CAM 更糟糕。這是因為 SEAM [41] 通過應用自我監督訓練來學習覆蓋目標物體的全部範圍, 這很容易被目標物體的重合像素所誤導。與此同時, CAM 只捕捉到目標物體的最具辨識性的區域, 並不覆蓋較不具辨識性的部分, 例如重合類別。我們也可以在圖 4 中觀察到這一現象。

表 3. 地圖選擇策略的效果。使用不同地圖選擇策略的偽掩碼準確性在 PASCAL VOC 2012 訓練集上進行評估。

方法	無 精煉	有 CRF [26]	有 AffinityNet [2]
CAM [52] <sub>CVPR'16</sub>	48.0	-	58.1
SEAM [41] <sub>CVPR'20</sub>	55.4	56.8	63.6
ICD [32] <sub>CVPR'20</sub> *	59.9	62.2	-
SGAN [47] <sub>ACCESS'20</sub> *	62.8	-	-
我們的 EPS	69.4	71.4	71.6

表 4. 在 PASCAL VOC 2012 訓練集上評估的偽掩碼的準確性 (mIoU)。注意, \* 表示忽略低置信度像素; 其他方法使用所有像素進行評估。

## 5.2. 地圖選擇策略的效果

我們評估了我們的地圖選擇策略在減少顯著性地圖錯誤方面的有效性。我們將三種不同的地圖選擇策略與不使用地圖選擇模塊的基線進行比較。作為簡單策略, 前景地圖是所有物體定位地圖的聯集; 背景地圖等於背景類別的定位地圖 (即簡單策略)。接下來, 我們遵循簡單策略, 但有以下例外。幾個預定類別 (例如, 沙發、椅子和餐桌) 的定位地圖被分配到背景地圖 (即預定類別策略)。最後, 所提出的選擇方法利用定位地圖和顯著性地圖之間的重疊比率, 如第 3.2 節所述 (即我們的自適應策略)。

Table 3 顯示, 我們的自適應策略可以有效地處理顯著性地圖的系統偏差。簡單策略意味著在從定位地圖生成估計的顯著性地圖時沒有考慮偏差。在這種情況下, 偽掩碼的性能下降, 特別是在沙發、椅子或餐桌類別上。使用預定類別的性能顯示, 通過忽略顯著性地圖中缺失的類別可以減少偏差。然而, 由於需要人類觀察者的手動選擇, 這不太實用, 並且無法對每張圖像做出最佳決策。與此同時, 我們的自適應策略可以自動處理偏差, 並為給定的顯著性地圖做出更有效的決策。





图 5. PASCAL VOC 2012 分割結果的質性範例。(a) 輸入圖像, (b) 真實標記和 (c) 我們的 EPS。

### 5.3. 與最先進技術的比較

**偽掩碼的準確性。**我們通過聚合來自不同尺度圖像的預測結果來採用多尺度推理，這是 [2, 41] 中常用的做法。然後，我們通過將我們的 EPS 與基線 CAM [52] 和三種最先進的方法，即 SEAM [41]、ICD [13] 和 SGAN [47] 進行比較，來評估訓練集中的偽掩碼的準確性。在 WSSS 中，測量訓練集中的偽掩碼的準確性是一種常見的協議，因為訓練集的偽掩碼用於監督分割模型。Table 4 總結了偽掩碼的準確性，並表明我們的方法明顯優於所有現有方法，差距達到 7–21

**分割地圖的準確性。**先前的方法 [2, 13, 41] 生成偽掩碼並使用 CRF 後處理算法 [26] 或親和網絡 [2] 進行細化。與此同時，如 Table 4 所示，我們生成的偽掩碼足夠準確，因此我們在不進行任何額外的偽掩碼細化的情況下訓練分割網絡。我們在 Pascal VOC 2012 數據集上的四個分割網絡上廣泛評估並精確比較我們的方法與其他方法。

我們的方法在分割網絡方面表現顯著優於其他方法。表 5 報告顯示，我們的方法在相同的 VGG16 骨幹下比其他方法更準確。此外，我們在 VGG16 上的結果與基於更強大骨幹的其他現有方法 (i.e. 表 6 中的 ResNet101) 相當甚至更優。我們的方法也顯示出對現有方法的明顯改進。最後，表 6 顯示，我們的方法（在基於 ResNet101 的 DeepLab-V1 與顯著性圖下）在 PASCAL VOC 2012 數據集上達到了新的最先進性能（驗證集為 71.0，測試集為 71.8）。我

方法	分割	支持	驗證	測試
SEC [25] <sub>ECCV'16</sub>	V1	I.	50.7	51.7
AffinityNet [2] <sub>CVPR'18</sub>	V1	I.	58.4	60.5
ICD [13] <sub>CVPR'20</sub>	V1	I.	61.2	60.9
BES [32] <sub>ECCV'20</sub>	V1	I.	60.1	61.1
GAIN [28] <sub>CVPR'18</sub>	V1	I.+S.	55.3	56.8
MCOF [40] <sub>CVPR'18</sub>	V1	I.+S.	56.2	57.6
SSNet [48] <sub>ICCV'19</sub>	V1	I.+S.	57.1	58.6
DSRG [20] <sub>CVPR'18</sub>	V2	I.+S.	59.0	60.4
SeeNet [19] <sub>NeurIPS'18</sub>	V1	I.+S.	61.1	60.7
MDC [44] <sub>CVPR'18</sub>	V1	I.+S.	60.4	60.8
FickleNet [27] <sub>CVPR'18</sub>	V2	I.+S.	61.2	61.9
OAA [21] <sub>ICCV'19</sub>	V1	I.+S.	63.1	62.8
ICD [13] <sub>CVPR'20</sub>	V1	I.+S.	64.0	63.9
Multi-Est. [14] <sub>ECCV'20</sub>	V1	I.+S.	64.6	64.2
Split. & Merge. [50] <sub>ECCV'20</sub>	V2	I.+S.	63.7	64.5
SGAN [47] <sub>ACCESS'20</sub>	V2	I.+S.	64.2	65.0
我們的 EPS	V1	I.+S.	66.6	67.9
	V2	I.+S.	67.0	67.3

表 5. 在 PASCAL VOC 2012 上的分割結果 (mIoU)。所有結果均基於 VGG16。最佳分數在所有實驗中以粗體顯示。

們強調，現有最先進模型所取得的增益約為 1%。同時，我們的方法比之前的最佳記錄高出超過 3% 的增益。圖 5 可視化了我們在 PASCAL VOC 2012 上的分割結果的質量示例。這些結果證實了我們的方法提供了準確的邊界並成功解決了共現問題。

在表 7 中，我們進一步在 COCO 2014 數據集中評估了我們的方法。我們使用基於 VGG16 的 DeepLab-V2 作為分割網絡來與 SGAN [47] 進行比

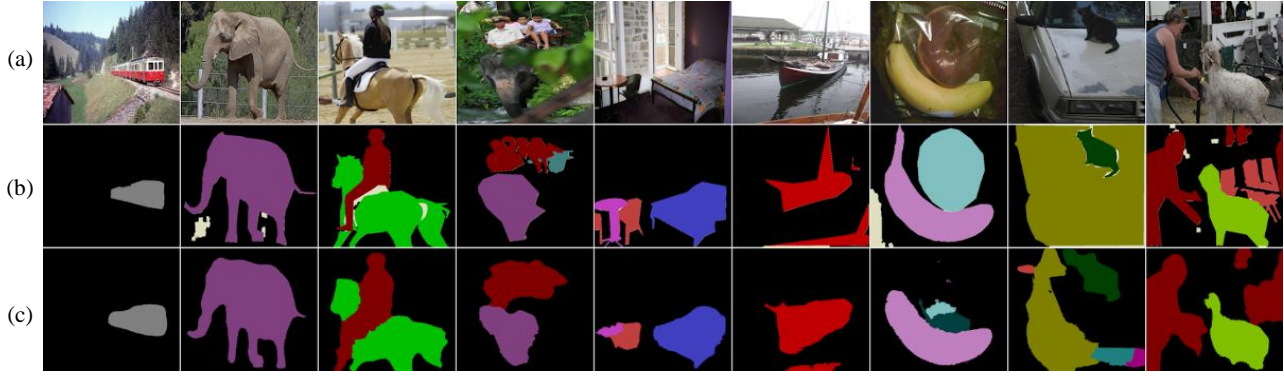


圖 6. MS COCO 2014 上分割結果的質性範例。(a) 輸入圖像, (b) 真實值和 (c) 我們的 EPS。

方法	分割	支持	驗證	測試
ICD [13] <sub>CVPR'20</sub>	V1	I.	64.1	64.3
SC-CAM [5] <sub>CVPR'20</sub>	V1	I.	66.1	65.9
BES [32] <sub>ECCV'20</sub>	V2	I.	65.7	66.6
LIID [31] <sub>TPAMI'20</sub>	V2	I.	66.5	67.5
MCOF [40] <sub>CVPR'18</sub>	V1	I.+S.	60.3	61.2
SeeNet [19] <sub>NeurIPS'18</sub>	V1	I.+S.	63.1	62.8
DSRG [20] <sub>CVPR'18</sub>	V2	I.+S.	61.4	63.2
FickleNet [27] <sub>CVPR'18</sub>	V2	I.+S.	64.9	65.3
OAA [21] <sub>ICCV'19</sub>	V1	I.+S.	65.2	66.4
Multi-Est. [14] <sub>ECCV'19</sub>	V1	I.+S.	67.2	66.7
MCIS [38] <sub>ECCV'20</sub>	V1	I.+S.	66.2	66.9
SGAN [47] <sub>ACCESS'20</sub>	V2	I.+S.	67.1	67.2
ICD [13] <sub>CVPR'20</sub>	V1	I.+S.	67.8	68.0
我們的 EPS	V1	I.+S.	71.0	71.8
	V2	I.+S.	70.9	70.8

表 6. PASCAL VOC 2012 上的分割結果 (mIoU)。所有結果均基於 ResNet101。

方法	分割	支持	驗證
SEC [25] <sub>ECCV'16</sub>	V1	I.	22.4
DSRG [20] <sub>CVPR'18</sub>	V2	I.+S.	26.0
ADL [9] <sub>TPAMI'20</sub>	V1	I.+S.	30.8
SGAN [47] <sub>ACCESS'20</sub>	V2	I.+S.	33.6
我們的 EPS	V2	I.+S.	35.7

表 7. 在 MS COCO 2014 上的分割結果 (mIoU)。所有結果均基於 VGG16。

較, 該模型是 COCO 數據集中的最先進 WSSS 模型。我們的方法在驗證集上達到了 35.7 mIoU, 比 SGAN [47] 高出 1.9%。因此, 我們在 COCO 2014

數據集中達到了新的最先進準確性。這些在兩個數據集上超越現有最先進技術的出色表現證實了我們方法的有效性; 通過充分利用定位圖和顯著性圖, 它成功地正確捕捉了目標對象的整體, 並彌補了現有模型的不足。圖 6 顯示了 COCO 2014 數據集上的分割結果的質量示例。我們的方法在少數對象出現而無遮擋時表現良好, 但在處理許多小對象時效果較差。更多示例和失敗案例在補充材料中提供。

**顯著性檢測模型的效果。**為了研究不同顯著性檢測模型的效果, 我們採用了三種顯著性模型; PFAN [51] (我們的默認), DSS [18] 用於 OAA [21] 和 ICD [13], 以及 USPS [34] (i.e., 無監督檢測模型)。在基於 Resnet101 的 DeepLab-V1 下的分割結果 (mIoU) 分別為 71.0/71.8 (PFAN), 70.0/70.1 (DSS), 和 68.8/69.9 (USPS) (驗證集和測試集)。這些分數支持我們的 EPS 使用任何三種不同的顯著性模型仍然比表 6 中的所有其他方法更準確。值得注意的是, 我們的 EPS 使用無監督顯著性模型優於所有使用監督顯著性模型的現有方法。

## 6. 結論

我們提出了一種新穎的弱監督分割框架, 即 顯式偽像素監督 (EPS)。受定位圖和顯著性圖之間互補關係的啟發, 我們的 EPS 從結合顯著性圖和定位圖的偽像素反饋中學習。由於我們的聯合訓練方案, 我們成功地補充了雙方的噪音或缺失信息。因此, 我們的 EPS 能夠捕捉精確的對象邊界並丟棄非目標對象的共現像素, 顯著提高了偽掩碼的質量。廣泛的



評估和各種案例研究證明了我們的 EPS 的有效性和出色的性能，為 PASCAL VOC 2012 和 MS COCO 2014 數據集上的 WSSS 提供了新的最先進準確性。

**致謝。**我們感謝 Duhyeon Bang 和 Junsuk Choe 的反饋。這項研究得到了韓國 MSIP 資助的 NRF 基礎科學研究計劃 (NRF-2019R1A2C2006123, 2020R1A4A1016619)，由 MSIT 資助的 IITP 資助 (2020-0-01361, 人工智能研究生院計劃 (延世大學))，以及由韓國政府資助的韓國醫療設備開發基金資助 (項目編號: 202011D06)。

## 参考文献

- [1] Jiwoon Ahn, Sunghyun Cho, and Suha Kwak. Weakly supervised learning of instance segmentation with inter-pixel relations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2209–2218, 2019. 2
- [2] Jiwoon Ahn and Suha Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4981–4990, 2018. 2, 6, 7
- [3] Nikita Araslanov and Stefan Roth. Single-stage semantic segmentation from image labels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4253–4262, 2020. 2
- [4] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1209–1218, 2018. 5
- [5] Yu-Ting Chang, Qiaosong Wang, Wei-Chih Hung, Robinson Piramuthu, Yi-Hsuan Tsai, and Ming-Hsuan Yang. Weakly-supervised semantic segmentation via sub-category exploration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8991–9000, 2020. 2, 8
- [6] Arslan Chaudhry, Puneet K. Dokania, and Philip H. S. Torr. Discovering class-specific pixels for weakly-supervised semantic segmentation. In *Proceedings of the British Machine Vision Conference*, 2017. 2
- [7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In *International Conference on Learning Representations*, 2015. 5
- [8] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017. 5
- [9] Junsuk Choe, Seungho Lee, and Hyunjung Shim. Attention-based dropout layer for weakly supervised single object localization and semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 1, 5, 8
- [10] Junsuk Choe, Seong Joon Oh, Seungho Lee, Sanghyuk Chun, Zeynep Akata, and Hyunjung Shim. Evaluating weakly supervised object localization methods right. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3133–3142, 2020. 2
- [11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009. 5
- [12] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, 2015. 5
- [13] Junsong Fan, Zhaoxiang Zhang, Chunfeng Song, and Tieniu Tan. Learning integral objects with intra-class discriminator for weakly-supervised semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4283–4292, 2020. 1, 3, 4, 6, 7, 8
- [14] Junsong Fan, Zhaoxiang Zhang, and Tieniu Tan. Employing multi-estimations for weakly-supervised semantic segmentation. In *Proceedings of the European Conference on Computer Vision*, 2020. 2, 7, 8
- [15] Junsong Fan, Zhaoxiang Zhang, Tieniu Tan, Chunfeng Song, and Jun Xiao. Cian: Cross-image affinity net for weakly supervised semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10762–10769, 2020. 2

- [16] Ruochen Fan, Qibin Hou, Ming-Ming Cheng, Gang Yu, Ralph R Martin, and Shi-Min Hu. Associating inter-image salient instances for weakly supervised semantic segmentation. In *Proceedings of the European Conference on Computer Vision*, pages 367–383, 2018. 2
- [17] Bharath Hariharan, Pablo Arbeláez, Lubomir Bourdev, Subhransu Maji, and Jitendra Malik. Semantic contours from inverse detectors. In *2011 International Conference on Computer Vision*, pages 991–998. IEEE, 2011. 5
- [18] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip HS Torr. Deeply supervised salient object detection with short connections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3203–3212, 2017. 2, 3, 8
- [19] Qibin Hou, PengTao Jiang, Yunchao Wei, and Ming-Ming Cheng. Self-erasing network for integral object attention. In *Advances in Neural Information Processing Systems*, pages 549–559, 2018. 7, 8
- [20] Zilong Huang, Xinggang Wang, Jiasi Wang, Wenyu Liu, and Jingdong Wang. Weakly-supervised semantic segmentation network with deep seeded region growing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7014–7023, 2018. 2, 5, 7, 8
- [21] Peng-Tao Jiang, Qibin Hou, Yang Cao, Ming-Ming Cheng, Yunchao Wei, and Hong-Kai Xiong. Integral object mining via online attention accumulation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2070–2079, 2019. 1, 3, 4, 7, 8
- [22] Anna Khoreva, Rodrigo Benenson, Jan Hosang, Matthias Hein, and Bernt Schiele. Simple does it: Weakly supervised instance and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 876–885, 2017. 1
- [23] Dahun Kim, Donghyeon Cho, Donggeun Yoo, and In So Kweon. Two-phase learning for weakly supervised object localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3534–3543, 2017. 1
- [24] Alexander Kolesnikov and Christoph Lampert. Improving weakly-supervised object localization by micro-annotation. In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference*, pages 92.1–92.12. BMVA Press, September 2016. 2
- [25] Alexander Kolesnikov and Christoph H Lampert. Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In *Proceedings of the European Conference on Computer Vision*, pages 695–711. Springer, 2016. 1, 2, 5, 7, 8
- [26] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in Neural Information Processing Systems*, pages 109–117, 2011. 6, 7
- [27] Jungbeom Lee, Eunji Kim, Sungmin Lee, Jangho Lee, and Sungroh Yoon. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5267–5276, 2019. 1, 2, 3, 7, 8
- [28] Kunpeng Li, Ziyang Wu, Kuan-Chuan Peng, Jan Ernst, and Yun Fu. Tell me where to look: Guided attention inference network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9215–9223, 2018. 1, 2, 3, 5, 7
- [29] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3159–3167, 2016. 1
- [30] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Proceedings of the European Conference on Computer Vision*, pages 740–755. Springer, 2014. 5
- [31] Yun Liu, Yu-Huan Wu, Pei-Song Wen, Yu-Jun Shi, Yu Qiu, and Ming-Ming Cheng. Leveraging instance-, image-and dataset-level information for weakly supervised instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2, 8
- [32] Chen Liyi, Wu Weiwei, Chenchen Fu, Xiao Han, and Yuntao Zhang. Weakly supervised semantic segmentation with boundary exploration. In *Proceedings of the European Conference on Computer Vision*, 2020. 1, 5, 6, 7, 8

- [33] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 891–898, 2014. 6
- [34] Tam Nguyen, Maximilian Dax, Chaithanya Kumar Mummadi, Nhung Ngo, Thi Hoai Phuong Nguyen, Zhongyu Lou, and Thomas Brox. Deepusps: Deep robust unsupervised saliency prediction via self-supervision. In *Advances in Neural Information Processing Systems*, pages 204–214, 2019. 2, 3, 8
- [35] Seong Joon Oh, Rodrigo Benenson, Anna Khoreva, Zeynep Akata, Mario Fritz, and Bernt Schiele. Exploiting saliency for object segmentation from image level labels. In *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 5038–5047. IEEE, 2017. 2, 5
- [36] Deepak Pathak, Philipp Krahenbuhl, and Trevor Darrell. Constrained convolutional neural networks for weakly supervised segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1796–1804, 2015. 1
- [37] Pedro O Pinheiro and Ronan Collobert. From image-level to pixel-level labeling with convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1713–1721, 2015. 1
- [38] Guolei Sun, Wenguan Wang, Jifeng Dai, and Luc Van Gool. Mining cross-image semantics for weakly supervised semantic segmentation. In *Proceedings of the European Conference on Computer Vision*, 2020. 2, 8
- [39] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 136–145, 2017. 2, 5
- [40] Xiang Wang, Shaodi You, Xi Li, and Huimin Ma. Weakly-supervised semantic segmentation by iteratively mining common object features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1354–1362, 2018. 2, 4, 7, 8
- [41] Yude Wang, Jie Zhang, Meina Kan, Shiguang Shan, and Xilin Chen. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12275–12284, 2020. 1, 3, 5, 6, 7
- [42] Yunchao Wei, Jiashi Feng, Xiaodan Liang, Ming-Ming Cheng, Yao Zhao, and Shuicheng Yan. Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1568–1576, 2017. 2, 3
- [43] Yunchao Wei, Xiaodan Liang, Yunpeng Chen, Xiaohui Shen, Ming-Ming Cheng, Jiashi Feng, Yao Zhao, and Shuicheng Yan. Stc: A simple to complex framework for weakly-supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11):2314–2320, 2016. 2, 4
- [44] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S Huang. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7268–7277, 2018. 2, 7
- [45] Zifeng Wu, Chunhua Shen, and Anton Van Den Hengel. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition*, 90:119–133, 2019. 5
- [46] Huaxin Xiao, Jiashi Feng, Yunchao Wei, Maojun Zhang, and Shuicheng Yan. Deep salient object detection with dense connections and distraction diagnosis. *IEEE Transactions on Multimedia*, 20(12):3239–3251, 2018. 2
- [47] Qi Yao and Xiaojin Gong. Saliency guided self-attention network for weakly and semi-supervised semantic segmentation. *IEEE Access*, 8:14413–14423, 2020. 2, 4, 6, 7, 8
- [48] Yu Zeng, Yunzhi Zhuge, Huchuan Lu, and Lihe Zhang. Joint learning of saliency detection and weakly supervised semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7223–7233, 2019. 2, 7
- [49] Bingfeng Zhang, Jimin Xiao, Yunchao Wei, Mingjie Sun, and Kaizhu Huang. Reliability does matter: An end-to-end weakly supervised semantic segmentation



approach. In Proceedings of the AAAI Conference on Artificial Intelligence, pages 12765–12772. AAAI Press, 2020. [2](#)

- [50] Tianyi Zhang, Guosheng Lin, Weide Liu, Jianfei Cai, and Alex Kot. Splitting vs. merging: Mining object regions with discrepancy and intersection loss for weakly supervised semantic segmentation. In Proceedings of the European Conference on Computer Vision, 2020. [2](#), [7](#)
- [51] Ting Zhao and Xiangqian Wu. Pyramid feature attention network for saliency detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3085–3094, 2019. [1](#), [2](#), [3](#), [4](#), [5](#), [8](#)
- [52] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2921–2929, 2016. [1](#), [5](#), [6](#), [7](#)