WACV 2021

ClassMix: Segmentation-Based Data Augmentation for Semi-Supervised Learning

Viktor Olsson1,2, Wilhelm Tranheden1,2,* Juliano Pinto1, Lennart Svensson1*

- **Problem/Objective**
  - Common augmentations used in semi-supervised classification are less effective for semantic segmentation.
- **Contribution/Key Idea**
  - Propose novel data augmentation mechanism called ClassMix.

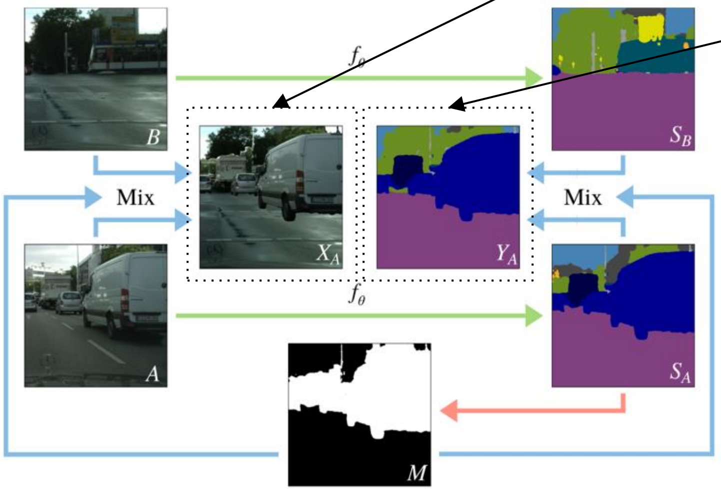
전유진

ClassMix: Segmentation-Based Data Augmentation for Semi-Supervised Learning

WACV 2021

3.1. ClassMix: Main Idea

ClassMix : Unlabeled image 2개로 새로운 augmented image와 이에 해당하는 artificial label를 만드는 기법.



1. 2개의 Unlabeled images (A, B) 를 샘플링함.
2. 위 두 이미지가 segmentation network, f_θ 를 통과한 결과 예측값인 S_A, S_B를 얻음.
3. S_A에서 존재하는 클래스들의 절반을 random 하게 선택하고, 선택된 클래스의 픽셀 값을 1로, 그리고 나머지 선택되지 않은 픽셀들은 모두 0으로 설정하여 이진마스크 M을 만듬.
4. 마스크 M을 사용하여 새로운 augmented image인 X_A를 만듬.
(마스크에서 1에 해당하는 픽셀들은 A로부터, 0에 해당하는 픽셀들은 B로부터 픽셀들을 가져와 새로운 이미지 X_A 생성).
5. 마찬가지로, 마스크 M을 사용하여 이미지 X_A에 상응하는 artificial label인 Y_A를 만듬.
(마스크에서 1에 해당하는 픽셀들은 S_A로부터, 0에 해당하는 픽셀들은 S_B로부터 픽셀들을 가져와 새로운 라벨 Y_A 생성).

전유진

3.2. ClassMix: Details

Algorithm 1 ClassMix algorithm

Require: Two unlabelled samples A and B, segmentation network $f_{\theta'}$.

$S_A \leftarrow f_{\theta'}(A)$

$S_B \leftarrow f_{\theta'}(B)$

$\tilde{S}_A \leftarrow \arg\max_{c'} S_A(i,j,c')$ ▷ Take pixel-wise argmax over classes.

$C \leftarrow$ Set of the different classes present in $\tilde{S}_A$

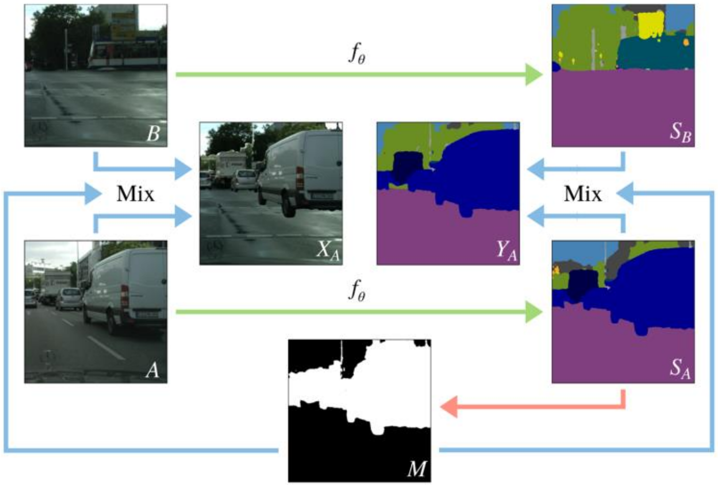
$c \leftarrow$ Randomly selected subset of C such that $|c| = |C|/2$

For all i, j: $M(i,j) = \begin{cases} 1, \text{ if } \tilde{S}_A(i,j) \in c \\ 0, \text{ otherwise} \end{cases}$ ▷ Create binary mask.

$X_A \leftarrow M \odot A + (1 - M) \odot B$ ▷ Mix images.

$Y_A \leftarrow M \odot S_A + (1 - M) \odot S_B$ ▷ Mix predictions.

return X_A, Y_A



-Mean-Teacher Framework.

1. 예측값인 S_A, S_B를 얻을 때, f_θ 대신 f_θ′ 모델 사용. (θ′ : 이전 θ 값들의 EMA)
2. X_A의 prediction은 f_θ 모델을 통해서 함.

-Pseudo-labelled Output.

Artifical label Y_A가 "argmaxed"된다.
즉, 각 픽셀에 대해 클래스 pmf가 가장 높은 클래스에 1, 나머지에는 0을 가진 원-핫 벡터로 변경된다.
-> 혼합 경계 예측이 특히나 불확실한데, 이 문제(라벨 오염 문제)를 완화시킴.

3.3. Loss and Training

Semantic segmentation network f_θ 를 아래 loss를 최소화함으로서 train 함.

$$L(\theta) = \mathbb{E}\left[\ell\Big(f_\theta(X_L), Y_L\Big) + \lambda\ell\Big(f_\theta(X_A), Y_A\Big)\right]$$

X_L : Image sampled uniformly at random from the labeled dataset.
Y_L : Corresponding gt semantic map.
X_A : Augmented image produced by ClassMix augmentation method.
Y_A : Corresponding artificial label produced by ClassMix augmentation method.
A, B : Input images are sampled uniformly at random from the unlabeled dataset.
λ : Hyper-parameter that controls the balance b/w supervised and unsupervised terms.
l (필기체 L) : Cross-entropy loss, averaged over all pixel positions in the semantic maps.

$$\ell(S, Y) = -\frac{1}{W \cdot H} \sum_{i=1}^{W} \sum_{j=1}^{H} \left(\sum_{c=1}^{C} Y(i,j,c) \cdot \log S(i,j,c) \right)$$

W, H : width and height of the images.
S(i,j,c), Y(i,j,c) : probabilities that the pixel in coordinates i,j belongs to class c, according to the prediction S and target Y, respectively.
위 loss에 SGD 방식으로 θ train함.
Batch : labeled data 50% + augmented data 50%.

Unsupervised weight λ starts close to zero initially.
This was accomplished by setting the value of λ for an augmented sample as the proportion of pixels in its artificial label where the probability of the most likely class is above a predetermined threshold τ . This results in a value between 0 and 1