Unsupervised domain adaptation based on the predictive uncertainty of models

23.04.12

이정민

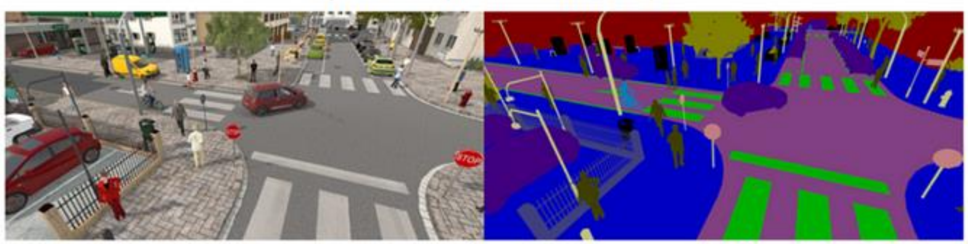
연구 배경

❖ Unsupervised Domain Adaptation(UDA)

- Target domain dataset에 labeled data가 없을 때

Domain Adaptation

Synthetic (source domain)



< Labeled dataset >

Real (target domain)

< Unlabeled dataset >

https://lhw0772.medium.com/study-da-domain-adaptation-%EC%95%8C%EC%95%84%EB%B3%B4%EA%B8%B0-%EA%B8%B0%EB%B3%B8%ED%8E%B8-4af4ab63f871

연구 배경

❖ 기존 UDA 방법론들의 한계점

- Domain divergence를 최소화하는 features 학습
- Feature extractor는 discriminator를 속이도록 학습(source와 target이 구분 안가게)

➡ **Class를 구분하는 정보를 고려하지 않음**

➡ 잘 구분되지 않는 target sample들은 **모델 불확실성**이 높을 것이다

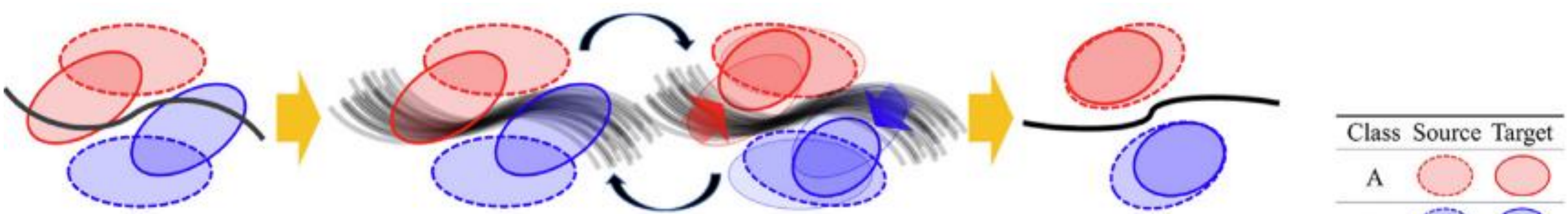
➡ Target sample들의 모델 불확실성이 낮아지도록 feature extractor를 학습하면, 서로 다른 domain에서 consistent한 feature를 생성할 수 있을 것이다

https://lhw0772.medium.com/study-da-domain-adaptation-%EC%95%8C%EC%95%84%EB%B3%B4%EA%B8%B0-%EA%B8%B0%EB%B3%B8%ED%8E%B8-4af4ab63f871

기여점

❖ **UDA method based on Model uncertainty(MUDA)**

- 모델 불확실성 기반의 UDA 방법론 제안
- Bayesian dropout을 사용하기 때문에 dropout을 사용하는 DNN모델에 모두 적용 가능
- Multi-source DA 문제에 적용 가능

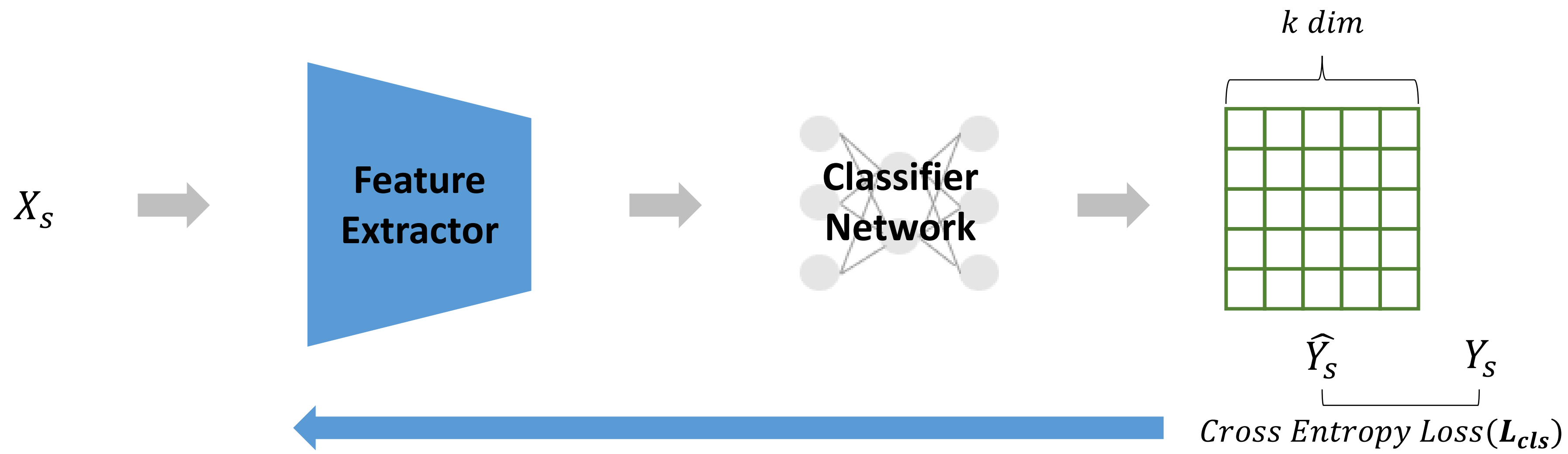


(a) Source only (b) Detect uncertainty & minimize uncertainty (c) Adapted

방법론

❖ Step1-Source domain training

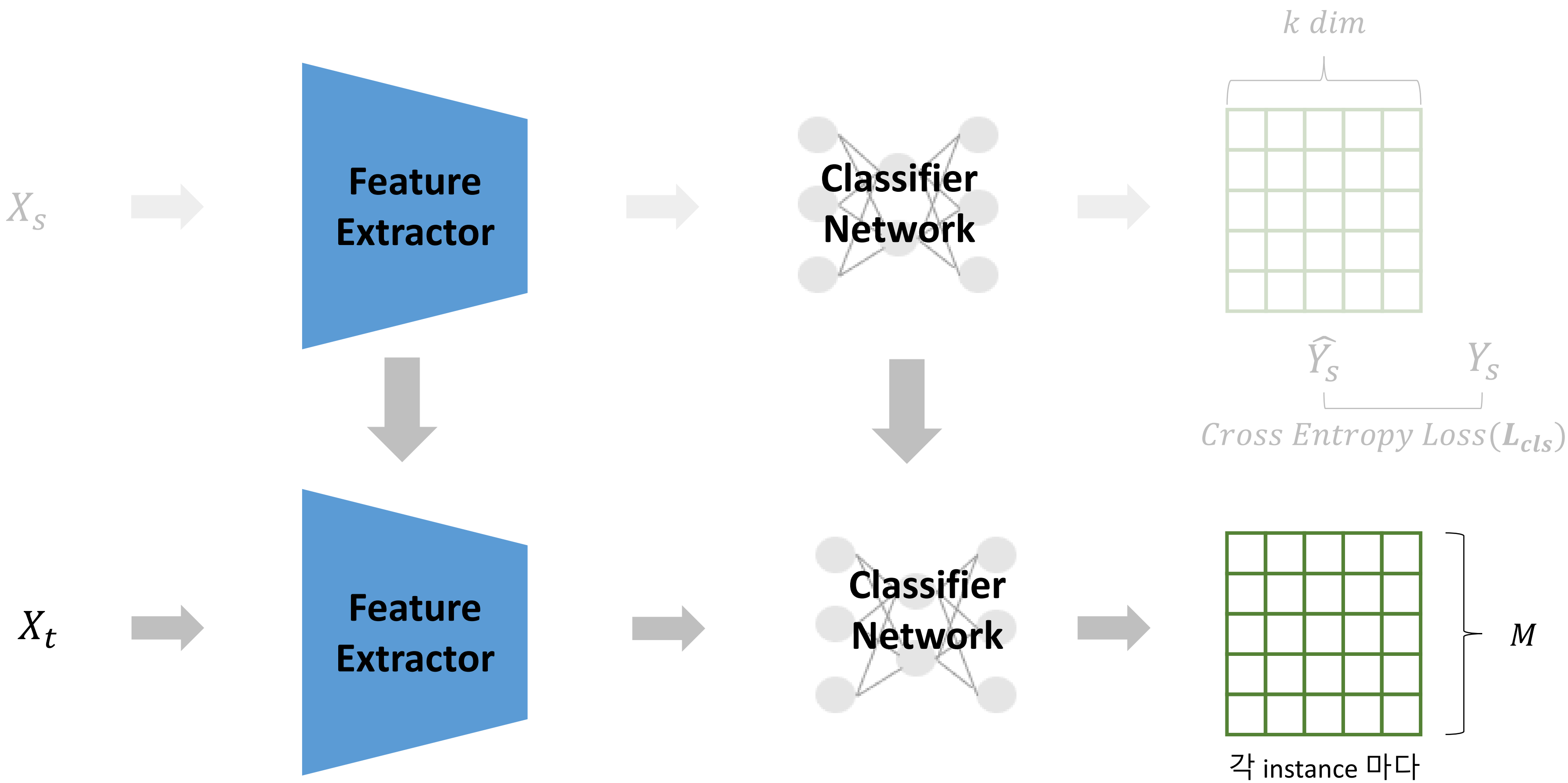
- $\{X_s, Y_s\}$으로 Source domain training (k: *number of class*)



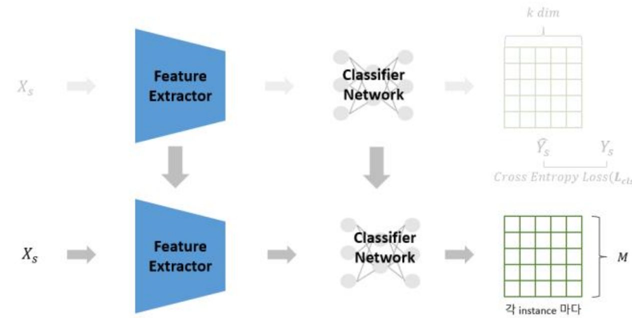
방법론

❖ **Step2-Target domain training**

- Step1에서 학습된 F, C 사용
- Step2에서는 F만 재학습



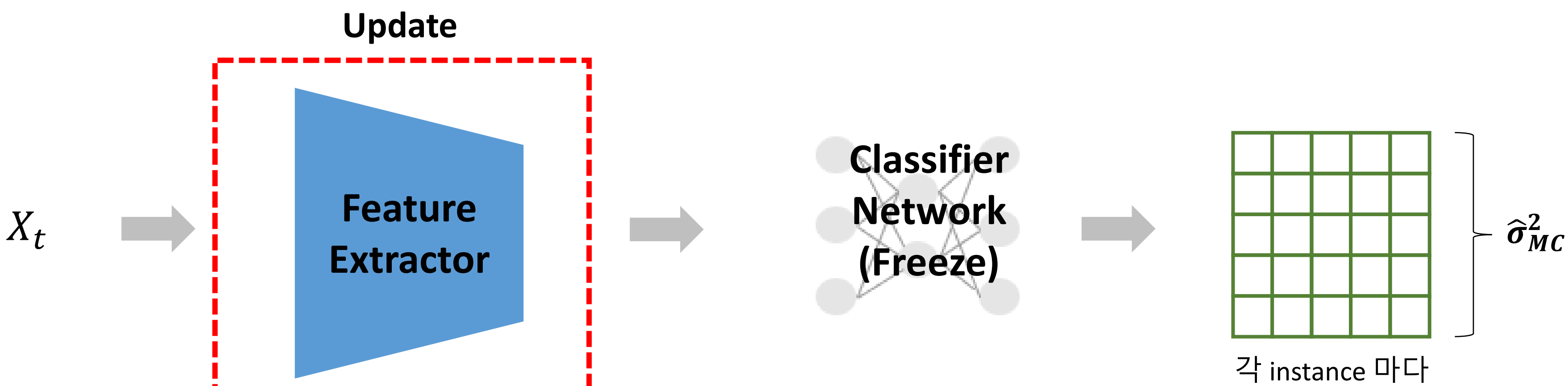
방법론



❖ Step2-Target domain training

- Step1에서 학습된 F, C 사용
- Step2에서는 F만 재학습

k dim

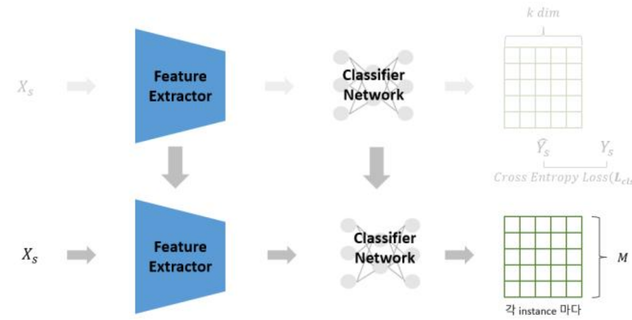


$$\hat{\sigma}^2_{MC}(x) = \mathrm{diag}\left(\frac{1}{M}\sum_{m=1}^{M}\hat{y}_m\hat{y}_m^{\top} - \bar{y}_{MC}\bar{y}_{MC}^{\top}\right),$$

$$\text{where } \bar{y}_{MC} = \frac{1}{M}\sum_{m=1}^{M}\hat{y}_m.$$

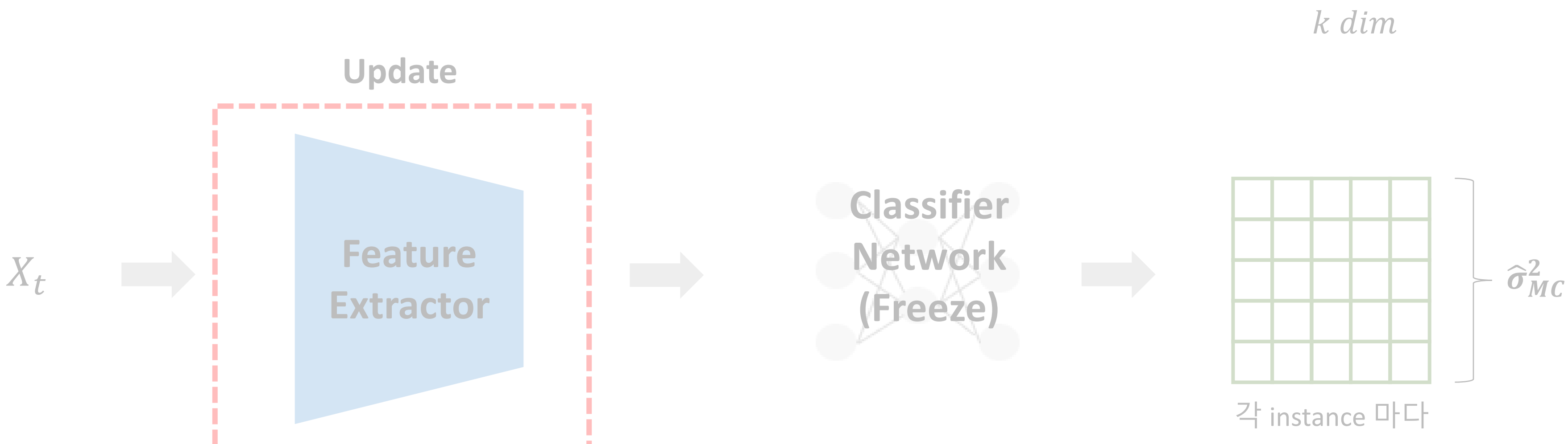
$$\mathcal{L}_{div}(\mathcal{D}_T) = \mathbb{E}_{x\sim\mathcal{D}_T}\|\hat{\sigma}_{MC}(x)\|$$

방법론



❖ Total Loss

- Classification Loss -> Classification Network 학습
- Classification Loss+Model Uncertainty Loss -> Feature Extractor 학습



$$\hat{\sigma}^2_{MC}(x) = \mathrm{diag}\left(\frac{1}{M}\sum_{m=1}^{M}\hat{y}_m\hat{y}_m^\top - \bar{y}_{MC}\bar{y}_{MC}^\top\right),$$

where $\bar{y}_{MC} = \frac{1}{M}\sum_{m=1}^{M}\hat{y}_m$.

$$\mathcal{L}_{div}(\mathcal{D}_T) = \mathbb{E}_{x\sim\mathcal{D}_T}\|\hat{\sigma}_{MC}(x)\|$$

$$\min_C \mathcal{L}_{cls}(\mathcal{D}_S)$$

$$\min_F \mathcal{L}_{cls}(\mathcal{D}_S) + \mathcal{L}_{div}(\mathcal{D}_T).$$