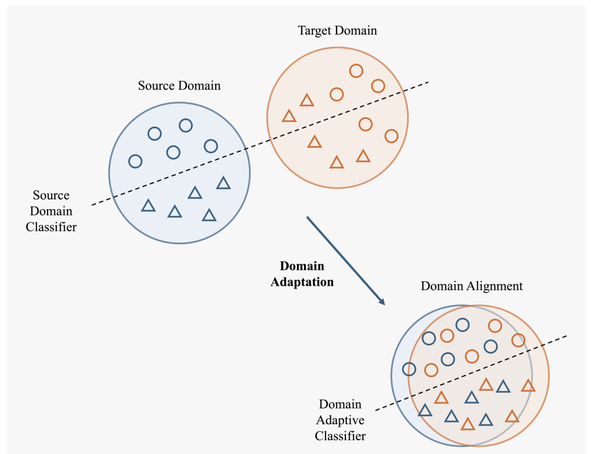
Selective Mutual Information Maximization for Source-free Domain Adaptation via Fusion of Foundation Models

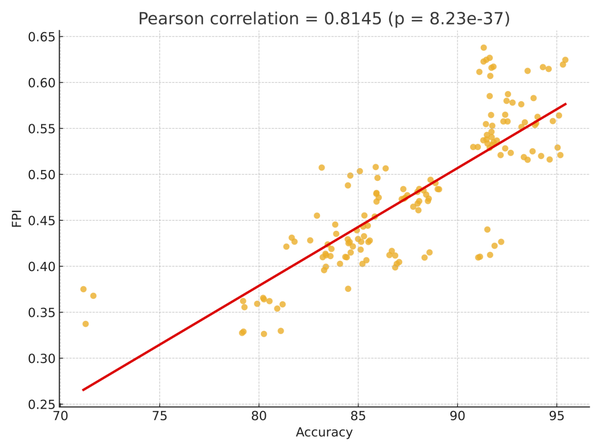
2025-1 AJOU SOFTCON

이 름 이휘수 지도교수 조현석

연구배경 및 관련연구

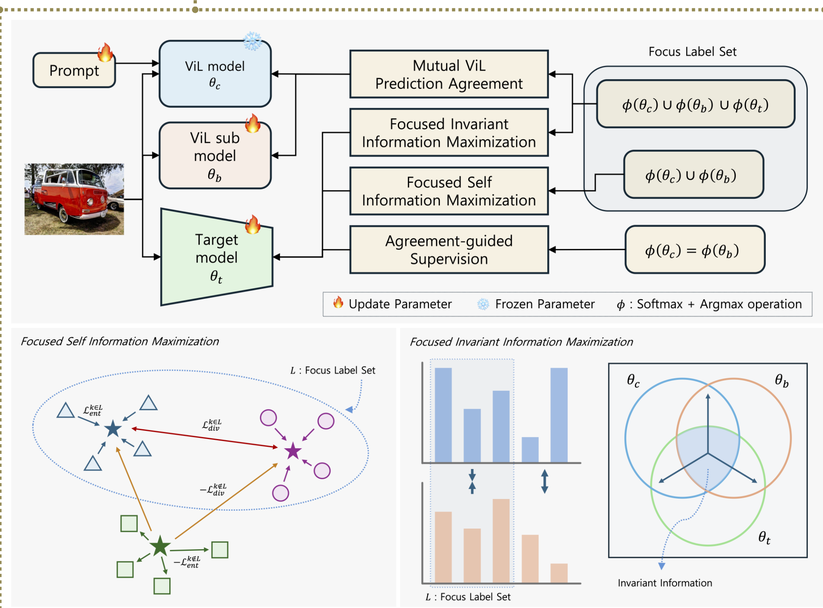
- Domain Adaptation (DA): 소스 도메인에서 학습한 모델을 타겟 도메인에 적용할 때, 성능 저하가 발생한다. DA는 이러한 문제를 해결하기 위해 소스 도메인과 타겟 도메인의 Feature Alignment를 수행하는 것을 말한다. 기존에는 소스 데이터와 Unlabeled 타겟 데이터를 사용하는 Unsupervised DA (UDA) 연구가 주를 이루었다.
- Source-free Domain Adaptation (SFDA): UDA 환경과 달리 현실적인 상황에서는 소스 데이터가 지적 재산권 문제와 컴퓨팅 리소스 부족 문제로 접근이 제한될 수 있다. 따라서 소스 데이터에 접근하지 않고, 오직 학습된 소스 모델과 타겟 이미지만을 이용하여 DA를 수행하는 SFDA 연구의 필요성이 제기되었다.
- Mutual Information (MI): MI는 두 확률변수의 의존성을 측정하는 척도로 기존 SFDA 연구에서는 MI를 최대화하여 예측의 불확실성을 최소화하거나 Inter model의 예측 일관성을 확보하는 형태로 활용되었다.
  - Contribution1: 기존의 MI 기반 방식은 배치 학습에서 불필요한 예측 다양성을 강제하거나, 신뢰도 낮은 예측에 대해서도 의존관계를 강제할 수 있다. 이에 따라 본 연구에서는 신뢰 기반의 class subset만을 대상으로 선택적으로 MI를 최대화하는 Decomposed Mutual Information (DMI) 최대화 방법을 제안한다.
- Multimodal Foundation Model (MFM): 최근 DIFO (2024, CVPR), ProDe (2025, ICLR) 등과 같이 SFDA에 MFM을 활용하는 연구사례가 등장하고 있다. 이는 MFM의 범용적 능력을 활용하여 SFDA의 성능을 극대화 하는 연구사례로 최신 SFDA 연구는 MFM을 사용하는 방향으로 발전하고 있다.
  - Contribution2: 본 연구에서는 Image-Text 유사성에 기반한 CLIP 모델 (2021, ICML)과 Text-Text 유사성에 기반한 BLIP 모델 (2023, ICML)의 서로 다른 표현 방식에 근거하여 두 모델과 타겟 모델을 상호 학습하여 효과적으로 SFDA를 달성하는 Framework를 제안한다.
  - Contribution3: 본 연구에서는 label이 없는 환경에서 MFM과 타겟 모델을 융합할 때, 사전적으로 Framework의 성능을 측정할 수 있는 지표인 Fusion Potential Index (FPI)를 정의하고, 실험적으로 FPI가 최종 성능과 높은 양의 상관관계를 보임을 입증하였다.





제안 기법

- Decomposed Mutual Information (DMI): 앞서 언급한 기존 MI의 문제를 해결하기 위해 수식 (1)과 같이 MI를 신뢰 집합 L에 기반하여 분해하고, L과 ~L의 정보량의 비율을 조절하기 위해 λ 값을 정의한다. 이는 L의 속한 class의 상호 정보는 극대화하고, L에 속하지 않은 class의 상호 정보는 억제하는 선택적 정보 극대화 방법이다. 수식 (2)는 앞서 정의한 DMI가 신뢰 집합 L의 크기에 따라 상한과 하한이 안정적으로 제한됨을 보여준다. 이는 DMI를 최대화할 때, L의 크기가 변하더라도 안정적으로 학습이 가능함을 의미한다.



[Model Training]

- Mutual ViL Prediction Agreement (MPA): MPA에서는 수식 (3)과 같이 CLIP과 BLIP 모델을 타겟 모델의 예측에 기반하여 Inter model DMI를 최대화한다. 이때, 수식 (4~5)와 같이 각 모델이 확신하는 정도를 반영하여 가중치를 부여한다. 이는 CLIP과 BLIP 모델을 task specific하게 customize하고, 상호 간의 불변하는 정보를 극대화함으로써 점진적으로 두 MFM이 합의에 이르도록 유도한다. (전체 framework 흐름과 L의 정의는 우측의 그림을 통해 확인할 수 있다.)
- Agreement-Guided Supervision (AGS): MPA를 통해 유도된 서로 합의된 예측은 서로 다른 표현 방식에 기반한 합의이므로 높은 신뢰성을 확보한다. 이에 따라 수식 (6)과 같이 서로 일치하는 예측으로 Pseudo-Label을 생성하여 Cross-entropy Loss로 타겟 모델을 학습한다. 이는 신뢰성 있는 예측에 기반하여 타겟 모델에 강력한 Supervision을 제공하여 빠르고 정확한 수렴을 유도한다.
- Focused Invariant Information Maximization (IIM): MPA에서는 Inter model DMI를 통해 MFM의 합의를 유도하였다면, IIM에서는 정렬된 MFM의 예측 분포에 기반하여 수식 (7)과 같이 Inter model DMI를 최대화하여 타겟 모델을 학습한다. 이를 통해 타겟 모델은 CLIP, BLIP, 타겟 모델 간의 상호 불변하는 정보를 극대화하여 타겟 모델의 특징 표현의 일관성과 신뢰성을 강화한다.
- Focused Self Information Maximization (SIM): SIM에서는 앞서 확보된 신뢰성 있는 Supervision과 상호 불변하는 정보에 기반하여 정렬된 타겟 모델의 예측 분포를 바탕으로, 타겟 모델이 스스로 예측의 불확실성을 줄이고, 일관성을 강화하는 자기 지도적인 정렬을 수행한다. 이는 수식 (8)과 같이 데이터 x와 모델의 예측 분포 $\theta_t(x)$ 간의 DMI를 최대화함으로써 달성된다.

[MFM Selection]

- Fusion Potential Index (FPI): CLIP과 BLIP 외에도 다양한 MFM이 해당 Framework를 통해 융합될 수 있다. 이때, FPI는 융합의 사전 판단 지표로써 활용될 수 있다. FPI는 수식 (9~11)과 같이 정의된다. 이는 서로 다른 모델의 예측이 일치할수록, 자기 확신이 높을수록 융합이 보완적 기능을 수행할 것이라는 직관에 기반하여 설계되었고, CLIP-B32, CLIP-B16, BLIP2, instBLIP, LLaVA의 다양한 조합으로 실험한 결과 우측 상단의 그림처럼 최종 성능과의 높은 양의 상관관계 $\rho = 0.81$ 을 보이며 유효한 지표로 작동함을 입증하였다.

$$I_d(X;Y) = I_{k\in L}(X;Y) - \lambda \cdot I_{k\notin L}(X;Y),\ \lambda = \begin{cases} \frac{\log|L|}{\log(K-|L|)} & \text{if } 0 \le |L| < K-1 \\ 0 & \text{if } K-1 \le |L| \le K \end{cases} \quad (1)$$

$$-\log|L| \le I_d(X;Y) \le \log|L| \quad (2)$$

$$\mathcal{L}_{\text{MPA}} = -\underset{v,\theta_b}{\text{argmin}}\ \mathbb{E}_{x_i\in\mathcal{X}_t}\Big\{\Big(\frac{w_c+w_b}{2}\Big)\cdot I_d(\theta_c(x_i,v),\theta_b(x_i)) + w_t\cdot I_d(\theta_c(x_i,v),\theta_t(x_i)) + w_t\cdot I_d(\theta_b(x_i),\theta_t(x_i))\Big\} \quad (3)$$

$$\text{Conf}_c = 1 - \frac{H(\delta(\theta_c))}{\log K},\ \text{Conf}_b = 1 - \frac{H(\delta(\theta_b))}{\log K},\ \text{Conf}_t = 1 - \frac{H(\delta(\theta_t))}{\log K} \quad (4)$$

$$w_i = \frac{\text{Conf}_i}{\text{Conf}_c + \text{Conf}_b + \text{Conf}_t}\ (i = c,\ b,\ t), \quad H(p) = -\sum_{i=1}^{K} p_i \log p_i \quad (5)$$

$$\mathcal{L}_{AGS} = -\mathbb{E}_{(x_i,\hat{y}_i)\in\mathcal{X}_t\times\hat{\mathcal{Y}}_t}\sum_{k=1}^{K} q_k \log \delta_k(\theta_t(x_i)),\ q_k = 1[k=\hat{y}_i] \quad (6)$$

$$\mathcal{L}_{IIM} = -\underset{\theta_t}{\text{argmin}}\mathbb{E}_{x_i\in\mathcal{X}_t}\{w_c\cdot I_d(\theta_c(x_i,v),\theta_t(x_i)) + w_b\cdot I_d(\theta_b(x_i),\theta_t(x_i))\} \quad (7)$$

$$\mathcal{L}_{SIM} = -\underset{\theta_t}{\text{argmin}}\mathbb{E}_{x_i\in\mathcal{X}_t} I_d(x_i,\theta_t(x_i)) \quad (8)$$

$$d(x) = \begin{cases} 1 & \text{if } \hat{y}_c(x) = \hat{y}_b(x) = \hat{y}_s(x) \\ 0.5 & \text{if only two match} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

$$\text{Agree}(\mathcal{X}) = \mathbb{E}_{x\sim\mathcal{X}}[d(x)] \quad (10)$$

$$\text{FPI} = \text{Agree} \times \text{Conf} \quad (11)$$

실험 결과 및 결론

전체 실험 결과는 우측의 QR 코드를 통해 확인할 수 있다.

- 실험: DA 표준 벤치마크인 Office31, OfficeHome, DomainNet, VisDA에서 실험을 진행하였다. 우측 하단의 표와 같이 OfficeHome에서 기존 최고 성능인 ProDe-V에 비해 6.1% 향상된 정확도를 보였고, 본 framework는 위의 모든 데이터셋에서 state-of-the-art (SOTA)의 성능을 달성함을 확인하였다. Office31, DomainNet, VisDA에서는 각각 기존 최고 성능에 비해 0.1%, 2.7%, 1.4%의 향상된 정확도를 기록하였다. (나머지 모든 실험 결과는 QR 코드를 통해 확인할 수 있다.)
- 결론: 본 연구는 기존의 MI의 문제를 해결하는 DMI를 제안하고, 서로 다른 표현 방식에 근거하여 MFM과 타겟 모델을 효과적으로 융합하는 SFDA framework를 제안하였다. 실험적 검증을 통해 다양한 데이터셋에서 SOTA의 성능을 달성함으로써 제안 기법의 우수성을 입증하였으며, FPI를 정의하여 융합의 유효성을 정량화함으로써, 현실적인 SFDA 환경에서 효과적으로 MFM을 활용할 수 있는 실질적이고, 확장 가능한 기반을 마련하였다.

Method	Venue	SF	M	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P	Avg.
Source	–	–	–	43.7	67.0	73.9	49.9	60.1	62.5	51.7	40.9	72.6	64.2	46.3	78.1	59.2
SHOT	ICML20	✓	✗	56.7	77.9	80.6	68.0	78.0	79.4	67.9	54.5	82.3	74.2	58.6	84.5	71.9
NRC	NIPS21	✓	✗	57.7	80.3	82.0	68.1	79.8	78.6	65.3	56.4	83.0	71.0	58.6	85.6	72.2
GKD	IROS21	✓	✗	56.5	78.2	81.8	68.7	78.9	79.1	67.6	54.8	82.6	74.4	58.5	84.8	72.2
AaD	NIPS22	✓	✗	59.3	79.3	82.1	68.9	79.8	79.5	67.2	57.4	83.1	72.1	58.5	85.4	72.7
AdaCon	CVPR22	✓	✗	47.2	75.1	75.5	60.7	73.3	73.2	60.2	45.2	76.6	65.6	48.3	79.1	65.0
CoWA	ICML22	✓	✗	56.9	78.4	81.0	69.1	80.0	79.9	67.7	57.2	82.4	72.8	60.5	84.5	72.5
ELR	ICLR23	✓	✗	58.4	78.7	81.5	69.2	79.5	79.3	66.3	58.0	82.6	73.4	59.8	85.1	72.6
PLUE	CVPR23	✓	✗	49.1	73.5	78.2	62.9	73.5	74.5	62.2	48.3	78.6	68.6	51.8	81.5	66.9
CPD	PR24	✓	✗	59.1	79.0	82.4	68.5	79.7	79.5	67.9	57.9	82.8	73.8	61.2	84.6	73.0
TPDS	IJCV24	✓	✗	59.3	80.3	82.1	70.6	79.4	80.9	69.8	56.8	82.1	74.5	61.2	85.3	73.5
DAPL-R	TNNLS23	✗	✓	54.1	84.3	84.8	74.4	83.7	85.0	74.5	54.6	84.8	75.2	54.7	83.8	74.5
PADCLIP-R	ICCV23	✗	✓	57.5	84.0	83.8	77.8	85.5	84.7	76.3	59.2	85.4	78.1	60.2	86.7	76.6
ADCLIP-R	ICCVW23	✗	✓	55.4	85.2	85.6	76.1	85.8	86.2	76.7	56.1	85.4	76.8	56.1	85.5	75.9
PDA-R	AAAI24	✗	✓	55.4	85.1	85.8	75.2	85.2	85.2	74.2	55.2	85.8	74.7	55.8	86.3	75.3
DAMP-R	CVPR24	✗	✓	59.7	88.5	86.8	76.6	88.9	87.0	76.3	59.6	87.1	77.0	61.0	89.9	78.2
DIFO-R	CVPR24	✓	✓	62.6	87.5	87.1	79.5	87.9	87.4	78.3	63.4	88.1	80.0	63.3	87.7	79.4
DIFO-V	CVPR24	✓	✓	70.6	90.6	88.8	82.5	90.6	88.8	80.9	70.1	88.9	83.4	70.5	91.2	83.1
ProDe-R	ICLR25	✓	✓	64.0	90.0	88.3	81.1	90.1	88.6	79.8	65.4	89.0	80.9	65.5	90.2	81.1
ProDe-V	ICLR25	✓	✓	72.7	92.3	90.5	82.5	91.5	90.7	82.5	72.5	90.8	83.0	72.6	92.2	84.5
Ours	–	✓	✓	**84.1**	**95.1**	**93.9**	**89.0**	**95.1**	**93.8**	**89.4**	**83.8**	**94.1**	**89.2**	**83.9**	**95.3**	**90.6**