

Domain Generalization by Solving Jigsaw Puzzles

Carlucci, F. M., D'Innocente, A., Bucci, S., Caputo, B., & Tommasi, T. (2019). Domain generalization by solving jigsaw puzzles. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2229-2238).

Introduction

Domain Generalization

- DG: 오직 source만을 사용해서, out-of-distribution data = unseen data에 대한 일반화 성능 향상을 목표
- 기존의 CNN은 image textures를 기반으로 학습을 진행[1]
- Image의 texture가 아닌 물체의 shape을 기반으로 학습을 진행할 수 있다면 domain 정보가 달라지더라도 좋은 예측 결과를 만들 수 있음
- 이를 위해 소개하는 논문은 Jigsaw puzzle pretext task 도입

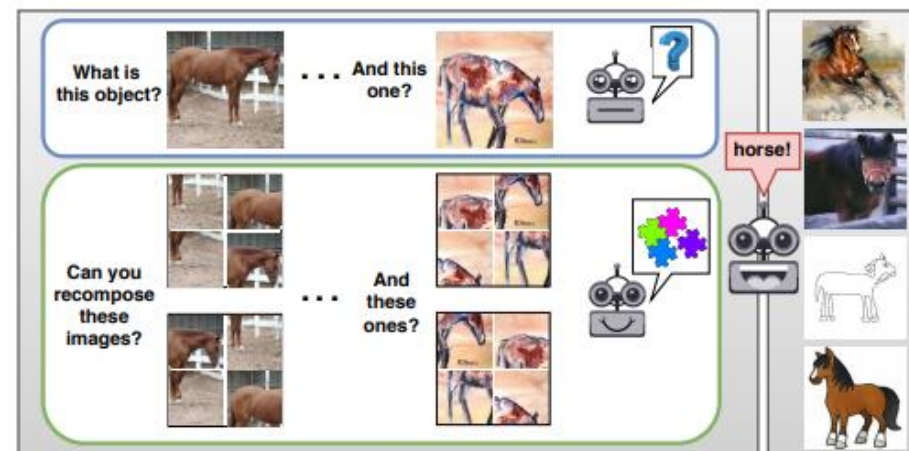
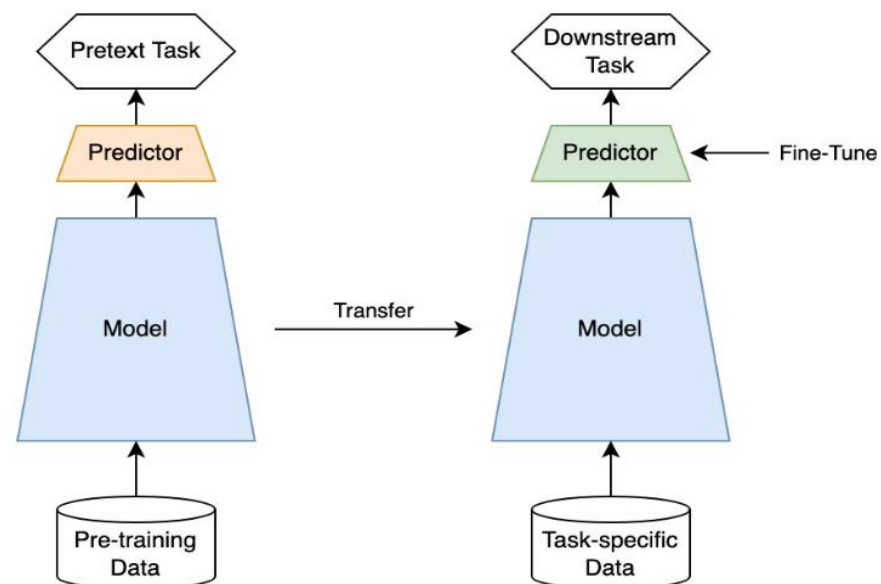


Figure 1. Recognizing objects across visual domains is a challenging task that requires high generalization abilities. Other tasks, based on intrinsic self-supervisory image signals, allow to capture natural invariances and regularities that can help to bridge across large style gaps. With JiGen we learn jointly to classify objects and solve jigsaw puzzles, showing that this supports generalization to new domains.

Related Work

Self Supervised Learning

- Self-supervised learning은 unlabeled dataset으로부터 좋은 representation을 얻고자 하는 학습 방식
- label 없이 input 내에서 self로 task를 정해서 supervision방식으로 모델을 학습
⇒ pretext task(=사용자가 새로운 문제를 정의하고 푸는 문제) problem
- pretext task를 학습한 모델은 downstream task에 transfer하여 사용



Related Work

Self Supervised Learning

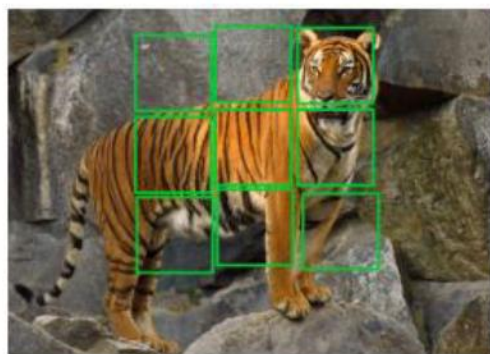
- ‘Task를 어떻게 정의하느냐’에 따라 크게 4가지의 방법으로 나뉨[NeurIPS2021의 [self-supervised learning 튜토리얼](#)]
1) Autoregressive generation 2) Masked generation 3) Innate relationship prediction 4) Hybrid self-prediction
- 그 중 Innate relationship prediction은 transformation을 하나의 샘플에 가했을 때도 **모델이 본질적인 정보(downstream)을 활용해 prediction**을 할 수 있도록 학습하는 방법
- augmentation이 어떻게 쓰였는지를 맞추는 task를 쓰게 되면 augmentation aware한 feature를 뽑을 수 있음
- Ex) 어떤 rotation이 적용되었는지 예측, patch기반의 jigsaw퍼즐을 풀, 시계열 데이터의 경우 segment shuffle로 순서를 classification



Related Work

Jigsaw Puzzle[2016 ECCV, Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles]

- 본 논문은 뒤죽박죽 섞어 놓고 각 patch(퍼즐 한 조각)의 특징만을 사용해서 원래의 배치로 돌아가기 위한 permutation을 예측하는 문제
- Jigsaw 퍼즐을 푸는 방식은 patch1, 2 같이 무늬, 색, 테두리, 바위 패턴 같은 shortcut들을 제한하고 representation/feature learning을 하는데 더욱 효과적
- 9개의 patch로 만들 수 있는 순열(permutation)은 $9! = 362,880$ 개 존재하지만 유사한 순열은 제거하면서 100개의 순열을 사용



Sample image



Extract 9 patches

Index (0~99)
61

Permutation
9, 5, 8, 3, 2, 4, 7, 1, 6

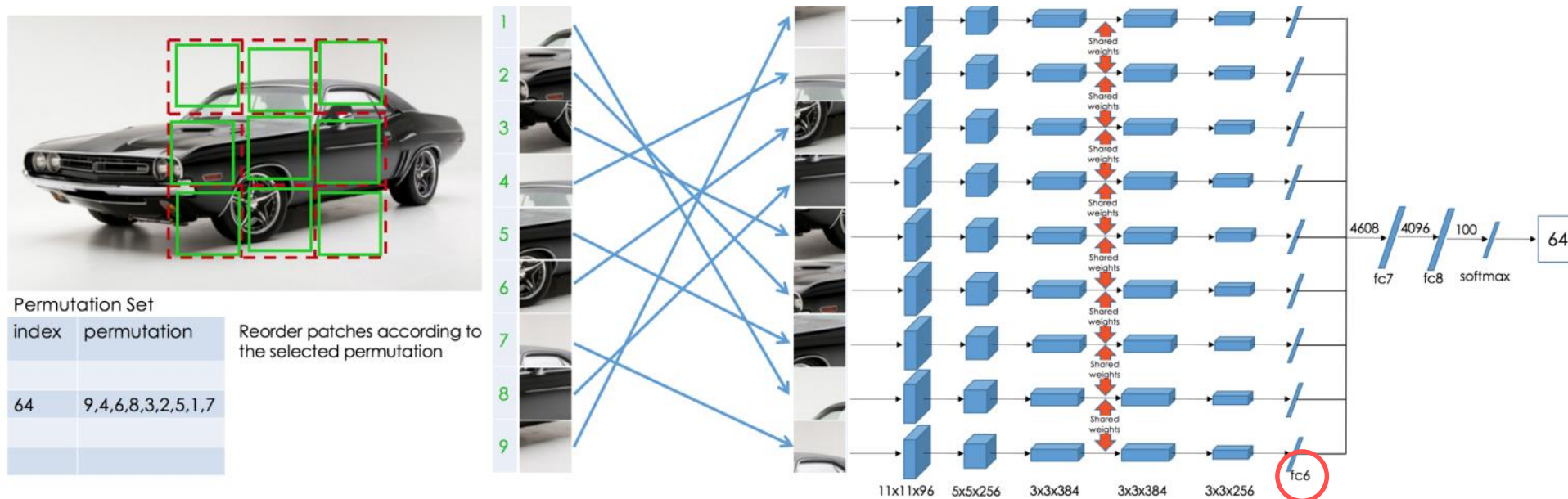


Permute 9 patches

Related Work

Jigsaw Puzzle[2016 ECCV, Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles]

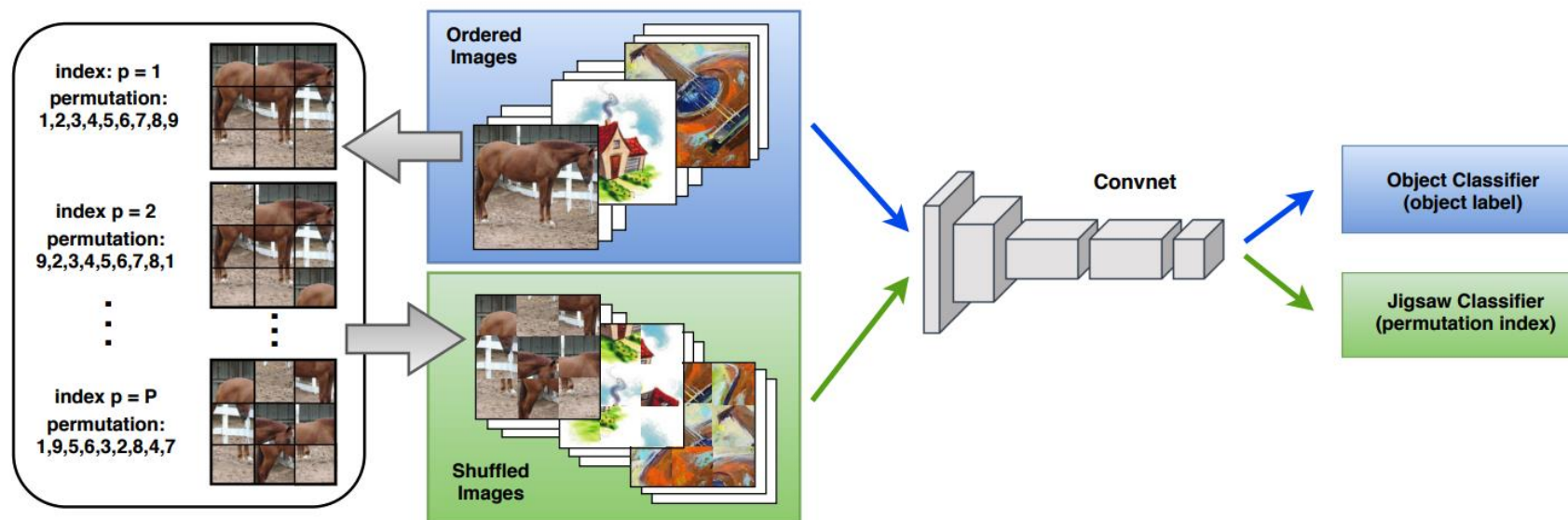
- 본 논문에서는 CFN(context free network) – Alexnet 제안(처음에는 각 타일의 특징들을 뽑아내고 마지막에 이를 이용해서 Jigsaw puzzle을 해결하는 구조)
 - 1개의 이미지를 9개의 patch로 무작위로 생성
 - 사전 정의된 순서 세트에서 하나에 따라서 재정렬 후 네트워크에 입력
 - fc6까지는 모두 shared weights를 사용하고, 정보의 흐름이 명시적으로 나뉘어져 있고 context는 fc7에서 사용



The JiGen Approach

Jigsaw Puzzle Based Generalization

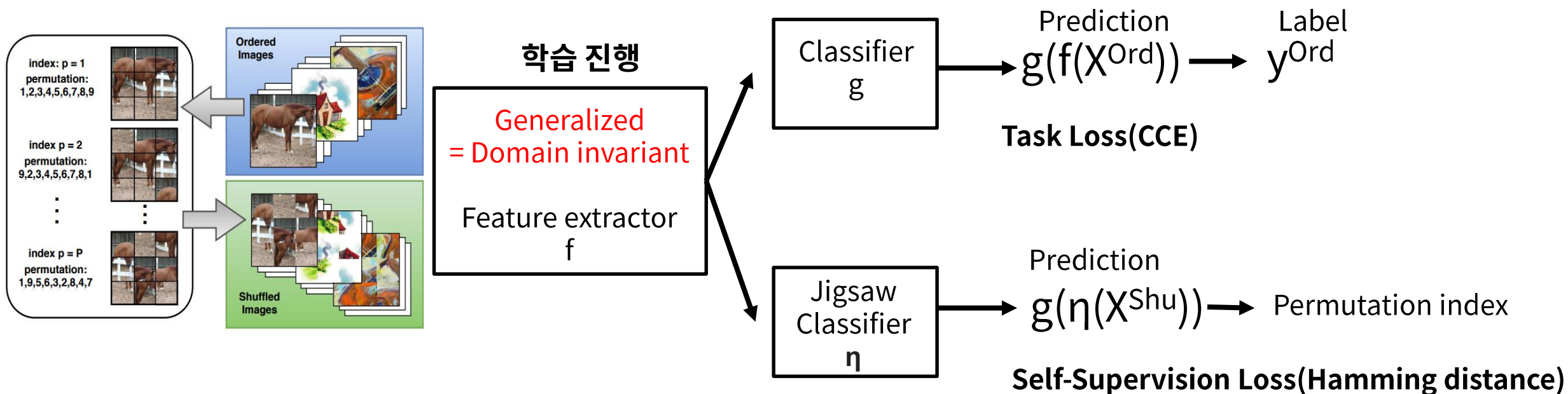
- Jigsaw 퍼즐을 푸는 self-supervision loss를 통해 domain invariant 특징(바나나 : 노란색, 길다, 살짝 휘었다,,,) 추출
- 즉, 기존의 classifier에서 DG를 위한 downstream을 수행하기 위해 Jigsaw classifier는 (generalization을 위한/ overfitting을 막는) **Network Regularizer 역할**



The JiGen Approach

Jigsaw Puzzle Based Generalization

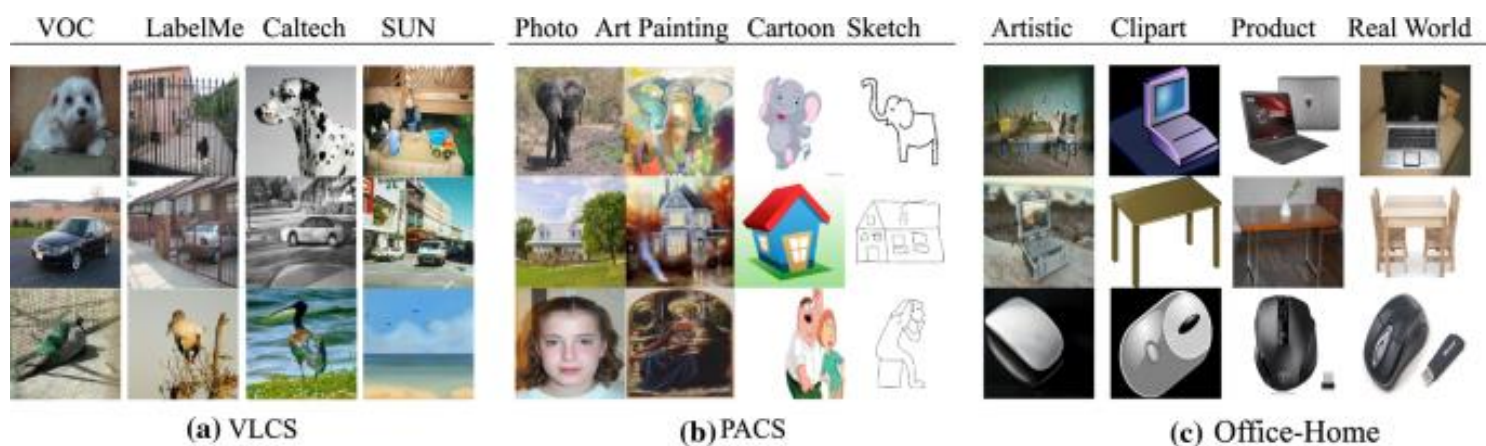
- Jigsaw 퍼즐을 푸는 self-supervision loss를 통해 domain invariant 특징(바나나 : 노란색, 길다, 살짝 휘었다,,,) 추출
- 즉, 기존의 classifier에서 DG를 위한 downstream을 수행하기 위해 Jigsaw classifier는 (generalization을 위한/ overfitting을 막는) **Network Regularizer 역할**



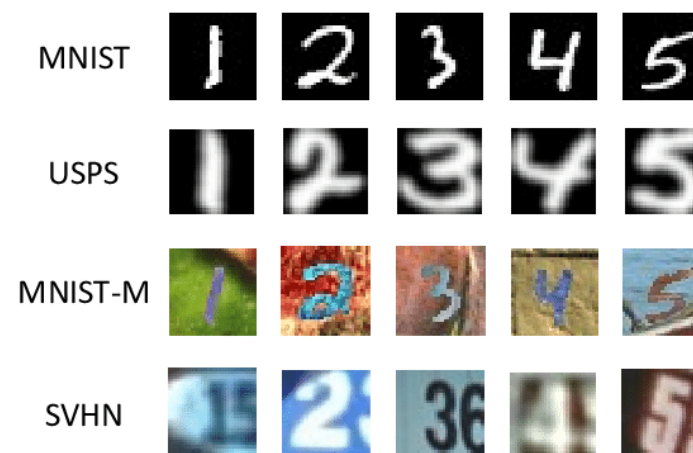
Experiments

Dataset

Multi-source DG



Single source DG



Experiments

Multi-source DG

- 3개의 backbone 네트워크로 다른 알고리즘과 비교 실험 결과 JiGen이 평균적으로 가장 좋은 성능을 보였다

	PACS	art_paint.	cartoon	sketches	photo	Avg.
CFN - Alexnet						
J-CFN-Finetune	47.23	62.18	58.03	70.18		59.41
J-CFN-Finetune++	51.14	58.83	54.85	73.44		59.57
C-CFN-Deep All	59.69	59.88	45.66	<u>85.42</u>		62.66
C-CFN-JiGen	60.68	60.55	55.66	82.68		64.89
Alexnet						
[26] Deep All	63.30	63.13	54.07	87.70		67.05
TF	62.86	66.97	57.51	89.50		69.21
Deep All	57.55	67.04	58.52	77.98		65.27
[28] DeepC	62.30	69.58	64.45	80.72		69.26
CIDDG	62.70	69.73	64.45	78.65		68.88
[25] Deep All	64.91	64.28	53.08	86.67		67.24
MLDG	66.23	66.88	58.96	88.00		70.01
[14] Deep All	64.44	<u>72.07</u>	58.07	87.50		70.52
D-SAM	63.87	70.70	64.66	85.55		71.20
Deep All	66.68	69.41	60.02	<u>89.98</u>		71.52
JiGen	67.63	71.71	65.18	89.00		73.38
Resnet-18						
[14] Deep All	77.87	<u>75.89</u>	69.27	95.19		79.55
D-SAM	77.33	72.43	77.83	95.30		80.72
Deep All	77.85	74.86	67.74	95.73		79.05
JiGen	79.42	75.25	71.35	96.03		80.51

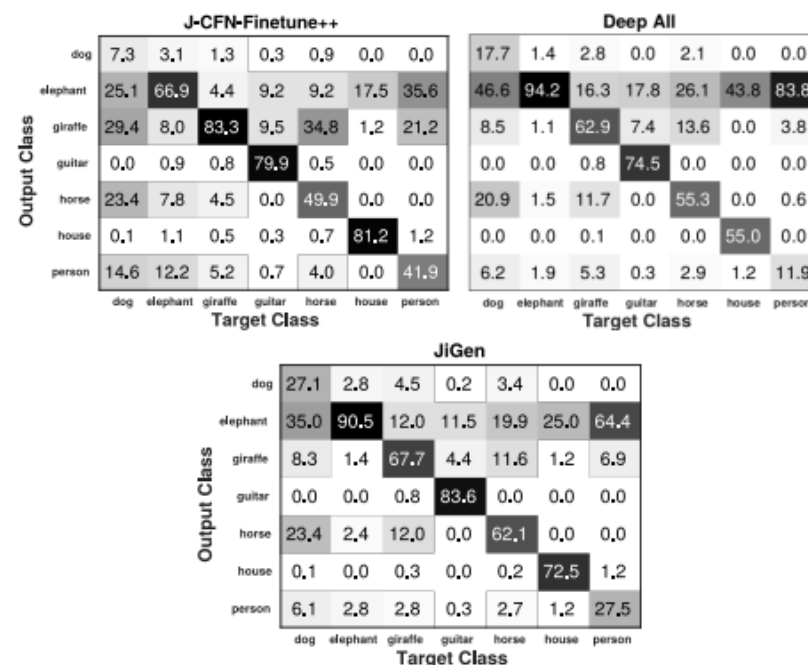


Figure 3. Confusion matrices on Alexnet-PACS DG setting, when sketches is used as target domain.

Experiments

Multi-source DG

- VLCS 와 Office Home dataset 또한 마찬가지로

	VLCS	Caltech	Labelme	Pascal	Sun	Avg.
Alexnet						
[28]	Deep All	85.73	61.28	62.71	59.33	67.26
	DeepC	87.47	62.60	63.97	61.51	68.89
	CIDDG	88.83	63.06	64.38	62.10	69.59
[33]	Deep All	86.10	55.60	59.10	54.60	63.85
	CCSA	92.30	62.10	67.10	59.10	70.15
[13]	Deep All	86.67	58.20	59.10	57.86	65.46
	SLRC	92.76	62.34	65.25	63.54	70.97
[26]	Deep All	93.40	62.11	68.41	64.16	72.02
	TF	93.63	63.49	69.99	61.32	72.11
[27]	MMD-AAE	94.40	62.60	67.70	64.40	72.28
[14]	Deep All	94.95	57.45	66.06	<u>65.87</u>	71.08
	D-SAM	91.75	56.95	58.59	60.84	67.03
	Deep All	<u>96.93</u>	59.18	<u>71.96</u>	62.57	72.66
	JiGen	96.93	60.90	70.62	64.30	73.19

Table 2. Domain Generalization results on VLCS. For details about number of runs, meaning of columns and use of bold/underline fonts, see Table 1.

	Office-Home	Art	Clipart	Product	Real-World	Avg.
Resnet-18						
[14]	Deep All	55.59	42.42	70.34	70.86	59.81
	D-SAM	58.03	44.37	69.22	71.45	60.77
	Deep All	52.15	45.86	70.86	<u>73.15</u>	60.51
	JiGen	53.04	47.51	71.47	72.79	61.20

Table 3. Domain Generalization results on Office-Home. For details about number of runs, meaning of columns and use of bold/underline fonts, see Table 1.

Experiments

Single-source DG

- Single DG에서는 MNIST-M을 사용했을 때는 전반적으로 더 안좋은 성능을 보이며 SVHN을 사용했을 경우 기존 방법보다 좋은 성능을 평균적으로 보이고 있음

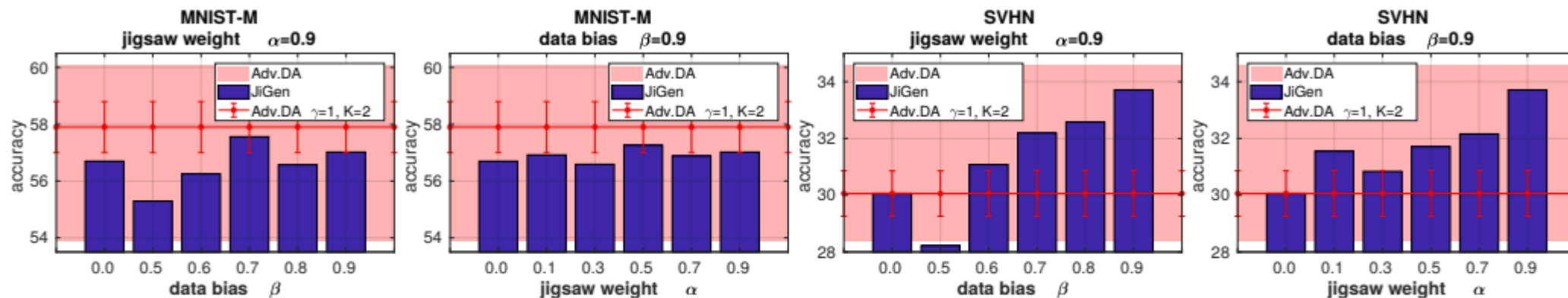


Figure 6. Single Source Domain Generalization experiments. We analyze the performance of JiGen in comparison with the method Adv.DA [46]. The shaded background area covers the overall range of results of Adv.DA obtained when changing the hyper-parameters of the method. The reference result of Adv.DA ($\gamma = 1, K = 2$) together with its standard deviation is indicated here by the horizontal red line. The blue histogram bars show the performance of JiGen when changing the jigsaw weight α and data bias β .

Conclusion

Conclusion

- Jigsaw puzzle 같이 인간의 눈에 변화가 보이는 task를 통해 image invariances를 학습하여 DG를 수행한 최초의 방법
- Pretrained model을 사용할 수 있으며 다양한 모델에 추가하여 보조적인 역할로 수행 가능한 장점이 존재

Future Work

- Image의 shape에 대한 focus 말고도 다른 regularities 를 도입한다면 보다 더 효과적으로 DG 가능할 수 있다
- 다른 모델과 접목하여 실제로 Jigsaw puzzle 방식이 DG에 효과적으로 작용하는지 test 필요