Heterogeneous Domain Generalization via Domain Mixup

Yufei Wang[1,2], Haoliang Li[2], and Alex C. Kot[2]

[1]University of Electronic Science and Technology of China, China

[2]Nanyang Technological University, Singapore

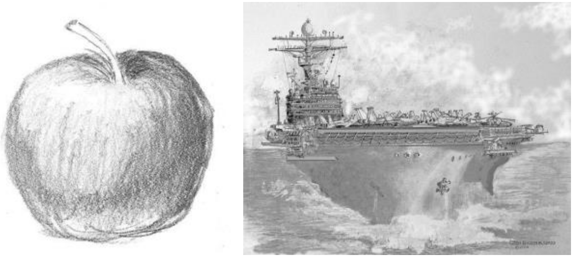
*Work done during an internship at the Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University (NTU)

*The code will be released in github

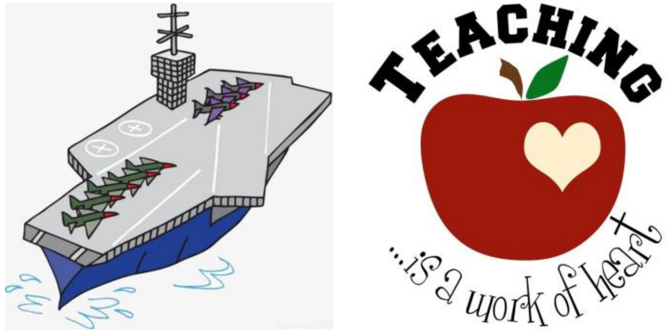
ICASSP 2020

Domain Generalization (DG)

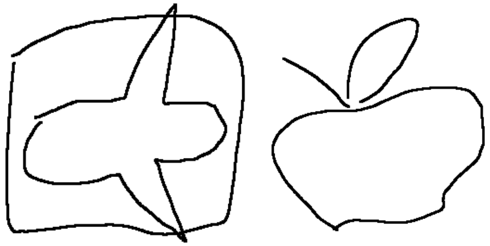
real



sketch

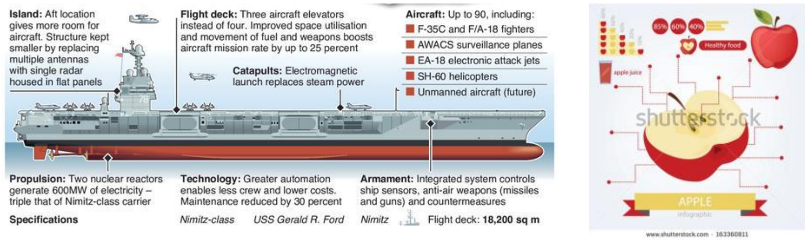


clipart



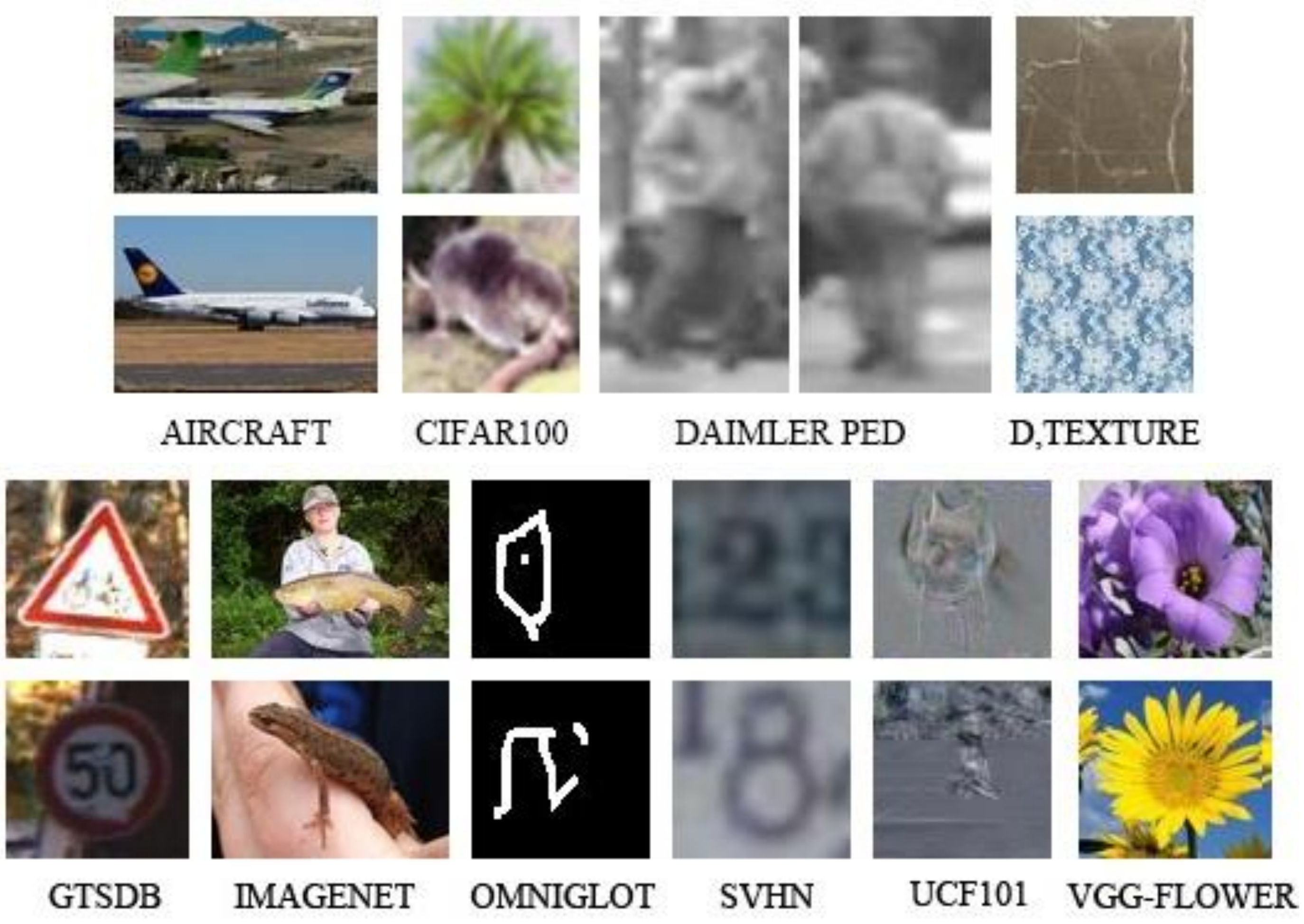
quick draw

painting

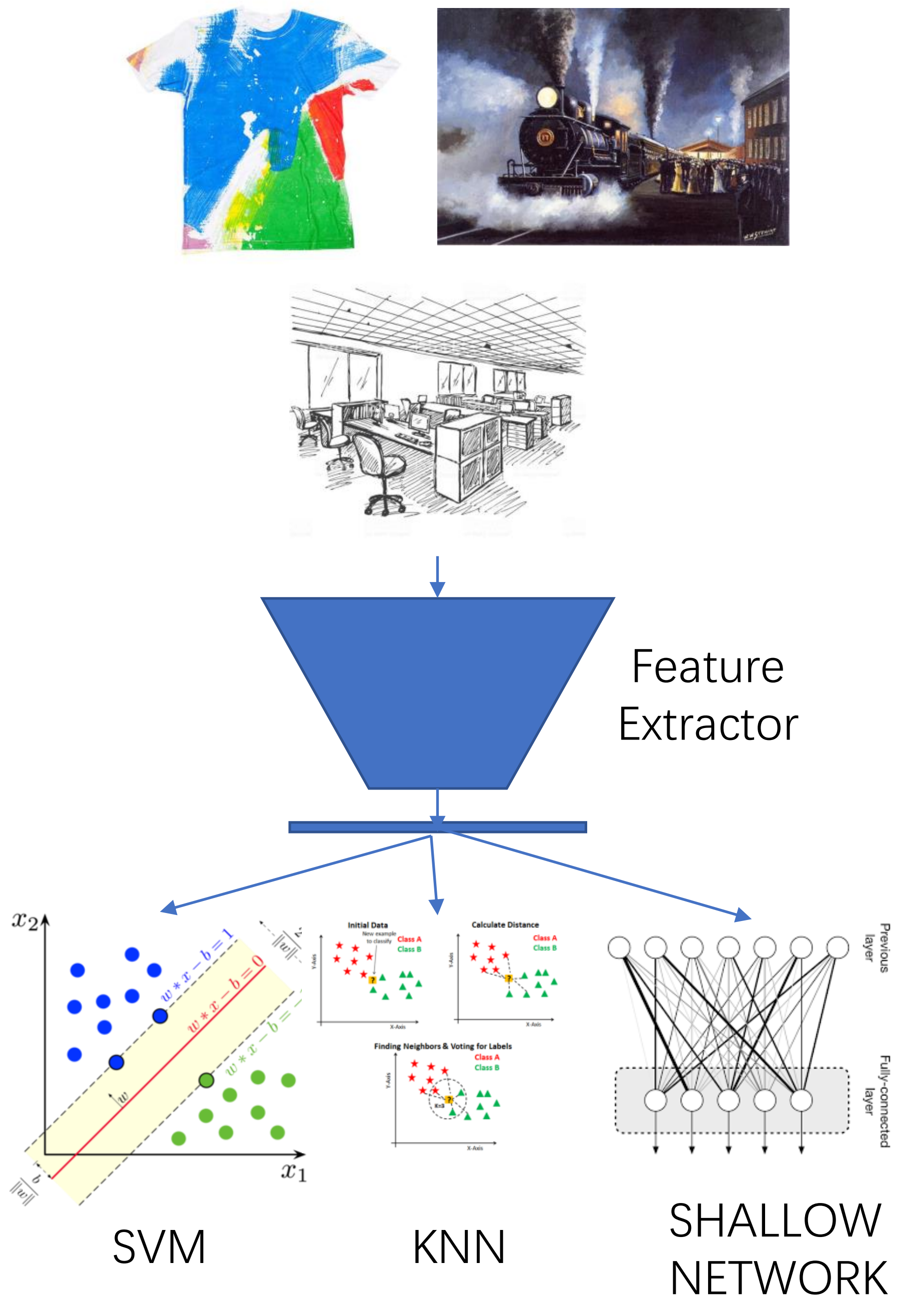


infograph

Heterogeneous DG



ICASSP 2020



Empirical Risk Minimization

- Minimize the average of the loss function l over the data distribution P, also known as the **expected risk:**

$$R(f) = \int l(f(x), y) \mathrm{d} P(x, y).$$

$$P_\delta(x, y) = \frac{1}{n} \sum_{i=1}^{n} \delta(x = x_i, y = y_i)$$

$$\mathrm{R}_\delta(\mathrm{f}) = \int l(\mathrm{f}(\mathrm{x}), \mathrm{y})\, d\mathrm{P}_\delta(\mathrm{x}, \mathrm{y}) = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} l(\mathrm{f}(\mathrm{x_i}), \mathrm{y_i})$$

Vicinal Risk Minimization

In the Vicinal Risk Minimization (VRM) principle (Chapelle et al., 2000), the distribution P is approximated by

- $P_{\nu}(\tilde{x}, \tilde{y}) = \frac{1}{n}\sum_{i=1}^{n} \nu(\tilde{x}, \tilde{y} | x_i, y_i)$

where ν is a vicinity distribution that measures the probability of finding the virtual feature-target pair $(\tilde{x}, \tilde{y})$ in the vicinity of the training feature-target pair (x_i, y_i).

MIXUP: a vicinal distribution

- Vanilla MIXUP

$$\mu(\tilde{x}, \tilde{y} | x_i^q, \hat{y}_i^q) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}_\lambda[\delta(\tilde{x} = \lambda x_i^q + (1-\lambda) x_j^q, \tilde{y} = \lambda \hat{y}_i^q + (1-\lambda) \hat{y}_j^q)]$$

- Cross-domain MIXUP

$$\mu_{\mathrm{HDG}}\left(\tilde{x}, \tilde{y} \middle| x_i^q, \widehat{y_i^q}\right) = \frac{1}{n_{sum}} \sum_{j=1}^{|\mathcal{D}_s|} \sum_{k=1}^{|D_j|} E_\lambda \left[\delta\left(\tilde{x} = \lambda \cdot x_i^q + (1-\lambda) \cdot x_k^j, \tilde{y} = \lambda \cdot \widehat{y_i^q} + (1-\lambda) \cdot \widehat{y_k^j}\right)\right]$$

Sampling strategies

1.MIXUP: A simple sampling strategy from the Cross-Domain MIXUP distribuiton

- $\tilde{x} = \lambda x_i^p + (1-\lambda) x_j^q$
- $\tilde{\mathrm{y}} = \lambda \widehat{\mathrm{y}_{\mathrm{i}}^{\mathrm{p}}} + (1-\lambda) \widehat{\mathrm{y}_{\mathrm{j}}^{\mathrm{q}}}$

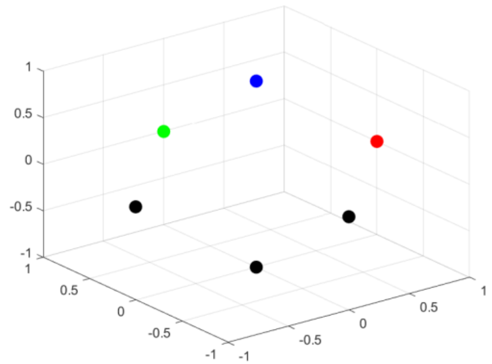
where λ~Beta(α,α) for α∈(0,∞)

2. **MIXALL:** mix all samples from |Ds| source domains in one iteration

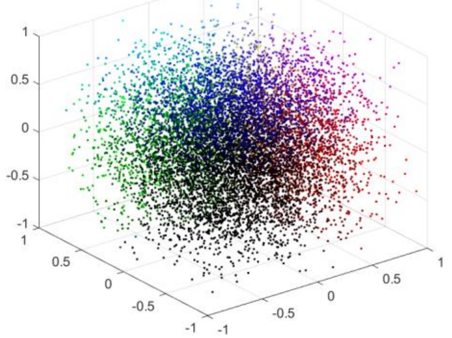
- $\tilde{x} = \sum_{q=1}^{|\mathcal{D}_s|} \lambda_q x_i^q$
- $\tilde{y} = \sum_{q=1}^{|\mathcal{D}_s|} \lambda_q \widehat{y_i^q}$

where $\lambda_{\mathrm{q}} = \frac{\mathrm{e}^{\mathrm{u_q}}}{\sum_{\mathrm{q}=1}^{|\mathcal{D}_s|} \mathrm{e}^{\mathrm{u_q}}}$ and random variable $\mathrm{u_q} \sim \mathrm{U}(0, \beta)$

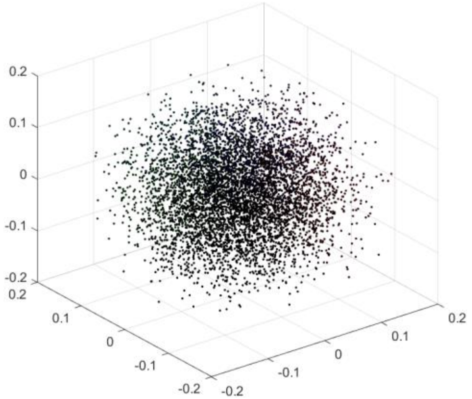
Multi-domain MIXUP: A toy example



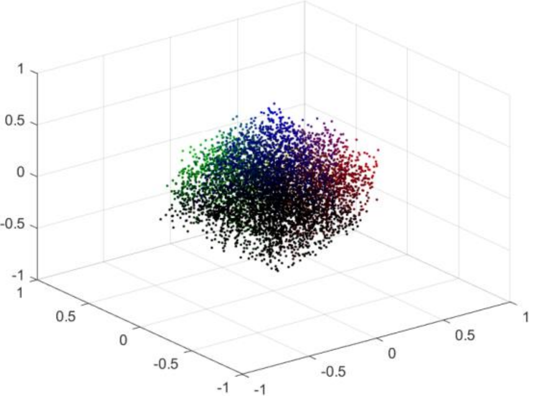
source domains



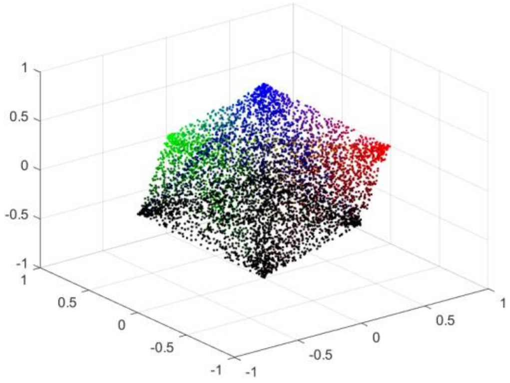
Random $\lambda \sim U(0,1)$



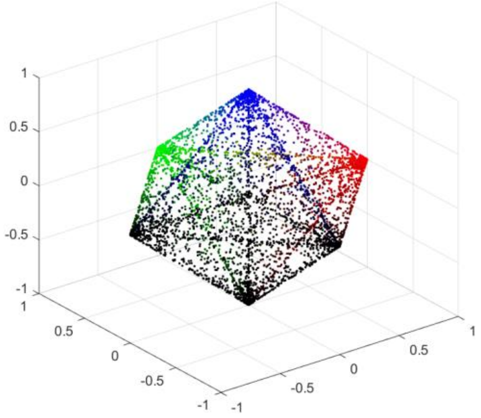
MIXALL with $\beta = 0$



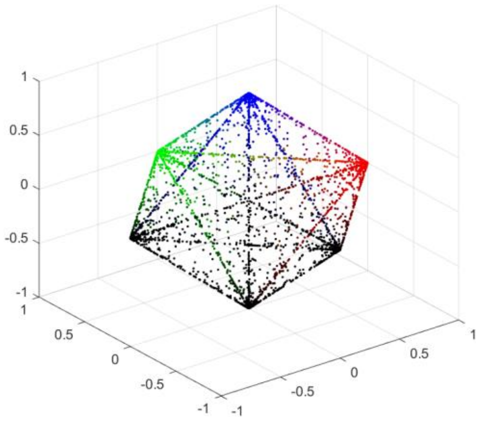
MIXALL with $\beta = 4$



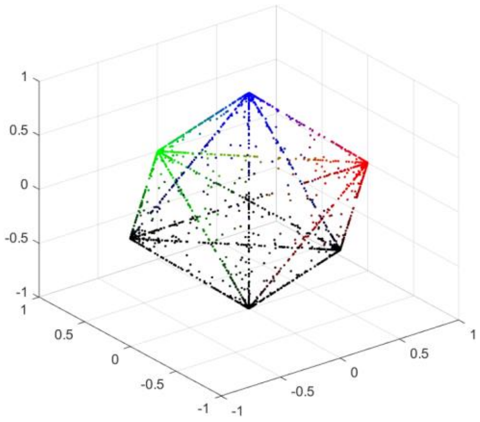
MIXALL with $\beta = 8$



MIXALL with $\beta = 16$



MIXALL with $\beta = 32$



MIXALL with $\beta = 64$

Result

Table 1. classification accuracy (%) and VD scores on target datasets in VD using SVM.

Target	SVM Classifier								
	Im.N. PT	CrossGrad [10]	MR [11]	MR-FL [14]	Reptile [12]	AGG [14]	FC [14]	MIXUP	MIX-ALL
Aircraft	16.62	19.92	20.91	18.18	19.62	19.56	20.94	**21.00**	20.61
D. Textures	41.70	36.54	32.34	35.69	37.39	36.49	38.88	**39.36**	39.04
VGG-Flowers	51.57	57.84	35.49	53.04	58.26	58.04	58.53	59.71	**61.37**
UCF101	44.93	45.80	47.34	48.10	49.85	46.98	50.82	51.74	**55.99**
Ave.	38.71	40.03	34.02	38.75	41.28	40.27	42.29	42.95	**44.25**
VD-Score	308	280	269	296	324	290	344	357	**400**

Table 2. classification accuracy (%) and VD scores on target datasets in VD using KNN.

Target	KNN Classifier								
	Im.N. PT	CrossGrad [10]	MR [11]	MR-FL [14]	Reptile [12]	AGG [14]	FC [14]	MIXUP	MIX-ALL
Aircraft	11.46	15.93	12.03	11.46	13.27	14.03	**16.01**	14.52	14.85
D. Textures	39.52	31.98	27.93	39.41	32.80	32.02	**34.92**	34.26	33.67
VGG-Flowers	41.08	48.00	23.63	39.51	45.80	45.98	47.04	48.24	**49.71**
UCF101	35.25	37.95	34.43	35.25	39.06	38.04	41.87	45.03	**46.41**
Ave.	31.83	33.47	24.51	31.41	32.73	32.52	34.96	35.51	**36.16**
VD-Score	215	188	144	215	201	189	236	259	**268**

Ablation Study

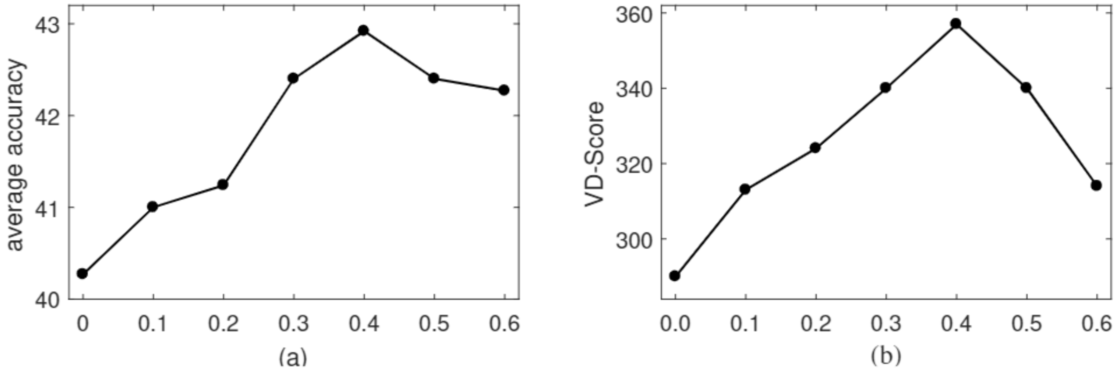


Fig. 2. The performance of MIXUP with different α which controls the strength of mixup.

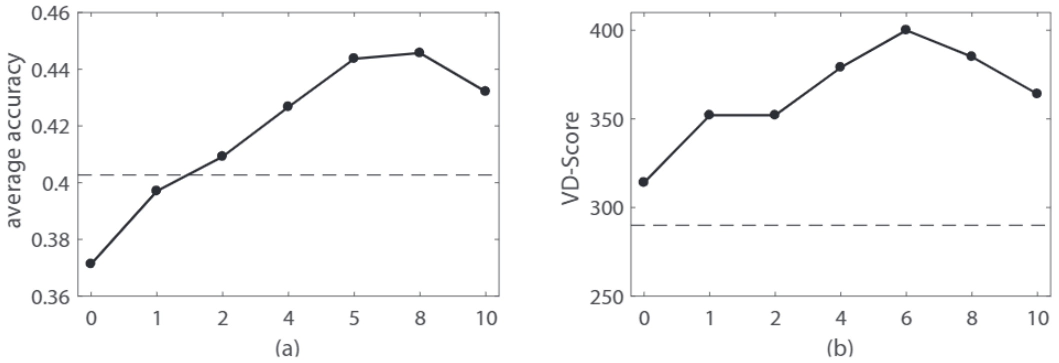


Fig. 3. The performance of MIX-ALL with different hyperparameter β which controls the strength of mixup. In this figure, the dotted line represents the baseline in Sec. 2.2

THANKS!