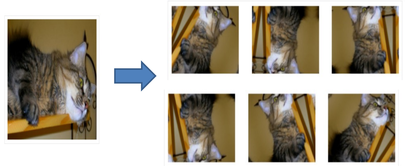
Cap2Aug: Caption guided image data Augmentation

Aniket Roy[1], Anshul Shah[1*], Ketul Shah[1*], Anirban Roy[2], Rama Chellappa[1]

[1]Johns Hopkins University [2]SRI International

Problem Description

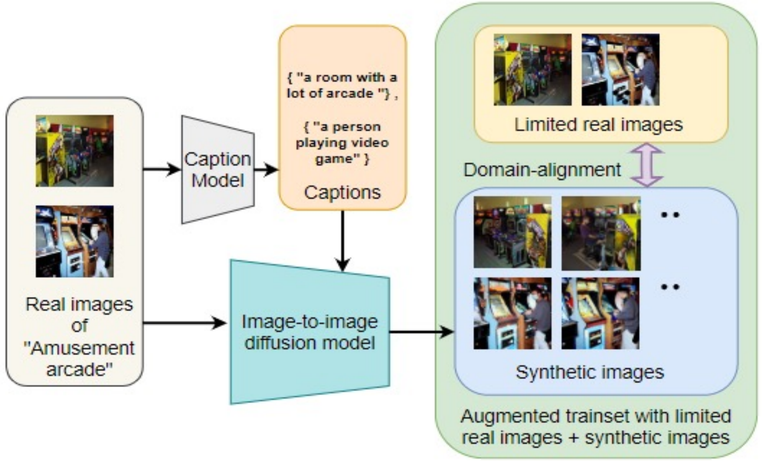
- Learning from small data is challenging.
- Potential solution is to use data augmentation.
- Standard data-augmentation methods do not provide semantic diversity.
- Idea: Can we use generative models to generate augmentations with semantic diversity?



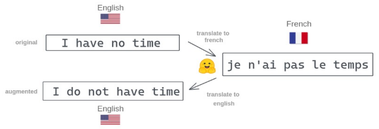
Standard data aug.

Generative data aug.

Method

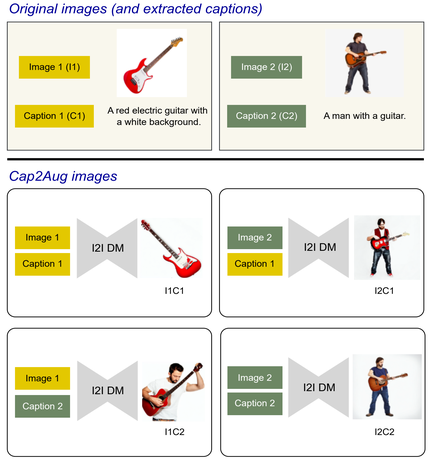


Cap2Aug



Back-translation

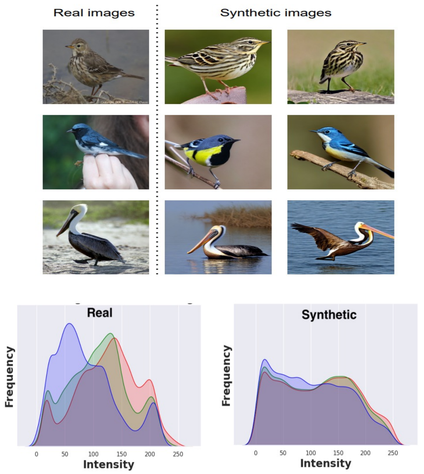
- Generate captions from the few-examples.
- Use these captions to generate augmented version of images using image-to-image diffusion model.



Potential Issue: Domain-gap

- Generated images might have domain-gap w.r.t real images.
- To mitigate this, we align the real and generated images using MMD loss.

$$\tilde{D}_k(P,Q) = \frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s} k(z_i^s, z_j^s) + \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t} k(z_i^t, z_j^t) - \frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j=1}^{n_t} k(z_i^s, z_j^t)$$

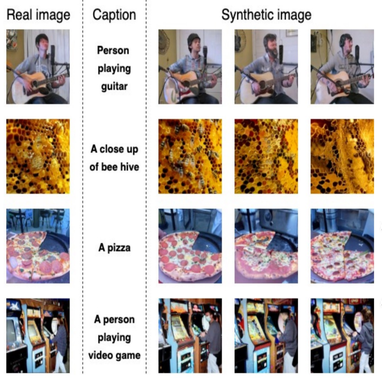


Task 1: Few-shot classification

	EuroSAT				SUN397				UCF101			
Shots	2	4	8	16	2	4	8	16	2	4	8	16
Tip [70]	61.68	65.32	67.95	70.50	62.70	64.15	65.62	66.85	64.74	66.46	68.68	70.58
CoOp [72]	61.50	70.18	76.73	82.53	59.48	63.47	65.52	69.26	64.09	67.03	71.92	75.71
Tip-F [70]	66.15	74.12	77.30	82.54	63.64	66.21	68.87	70.47	66.43	70.55	74.01	77.03
Ours	**67.03**	**77.37**	**77.50**	**83.64**	**64.61**	**67.47**	**68.93**	**70.90**	**68.57**	**71.76**	**74.12**	**77.24**
Δ	+0.88	+3.25	+0.20	+1.10	+0.97	+1.26	+0.06	+0.43	+2.14	+1.21	+0.11	+0.21

	OxfordPets				OxfordFlowers				FGVC			
Shots	2	4	8	16	2	4	8	16	2	4	8	16
Tip [70]	87.03	86.45	87.03	88.14	79.13	83.80	87.98	89.89	21.21	22.41	25.59	29.76
CoOp [72]	82.64	86.70	85.32	87.01	77.50	85.20	90.18	94.51	18.68	21.87	26.13	31.26
Tip-F [70]	87.03	87.54	88.09	89.70	82.30	85.83	90.51	**94.80**	23.19	24.80	29.21	**34.55**
Ours	**87.33**	**87.92**	**88.20**	**89.73**	**83.06**	**86.64**	**91.44**	94.51	**23.76**	**24.87**	**29.82**	34.38
Δ	+0.30	+0.38	+0.11	+0.03	+0.76	+0.81	+0.93	-0.29	+0.57	+0.07	+0.61	-0.17

	StanfordCars				Food101				DTD			
Shots	2	4	8	16	2	4	8	16	2	4	8	16
Tip [70]	57.93	61.45	62.90	66.77	77.52	77.54	77.76	77.83	49.47	53.96	58.63	60.93
CoOp [72]	58.28	62.62	68.43	73.36	72.49	73.33	71.82	74.67	45.15	53.49	59.97	63.58
Tip-F [70]	61.10	64.50	68.25	74.15	77.60	77.80	78.10	79.00	53.72	57.39	62.70	65.50
Ours	**61.25**	**64.70**	**69.15**	**74.80**	**77.66**	**77.89**	**78.47**	**79.05**	**54.50**	**59.28**	**63.47**	**66.13**
Δ	+0.15	+0.20	+0.90	+0.65	+0.06	+0.09	+0.37	+0.05	+0.78	+1.89	+0.77	+0.63



Method	**2-shot**	**4-shot**	**8-shot**	**16-shot**
Tip [70]	60.96	60.98	61.45	62.03
CoOp [72]	50.88	56.22	59.93	62.95
Tip-F [70]	61.69	62.52	64.00	65.51
TaskRes [66]	62.17	62.93	64.03	64.75
Ours	**62.72**	**63.50**	**64.95**	**66.32**
Δ	+0.55	+0.57	+0.92	+0.81