SaSPA: Advancing Fine-Grained Classification by Structure and Subject Preserving Augmentation

NeurIPS 2024

Computer Science MLDS School

**Eyal Michaeli
Ohad Fried**

What is FGVC?

- **Differences between classes** are usually minor, yet **important**

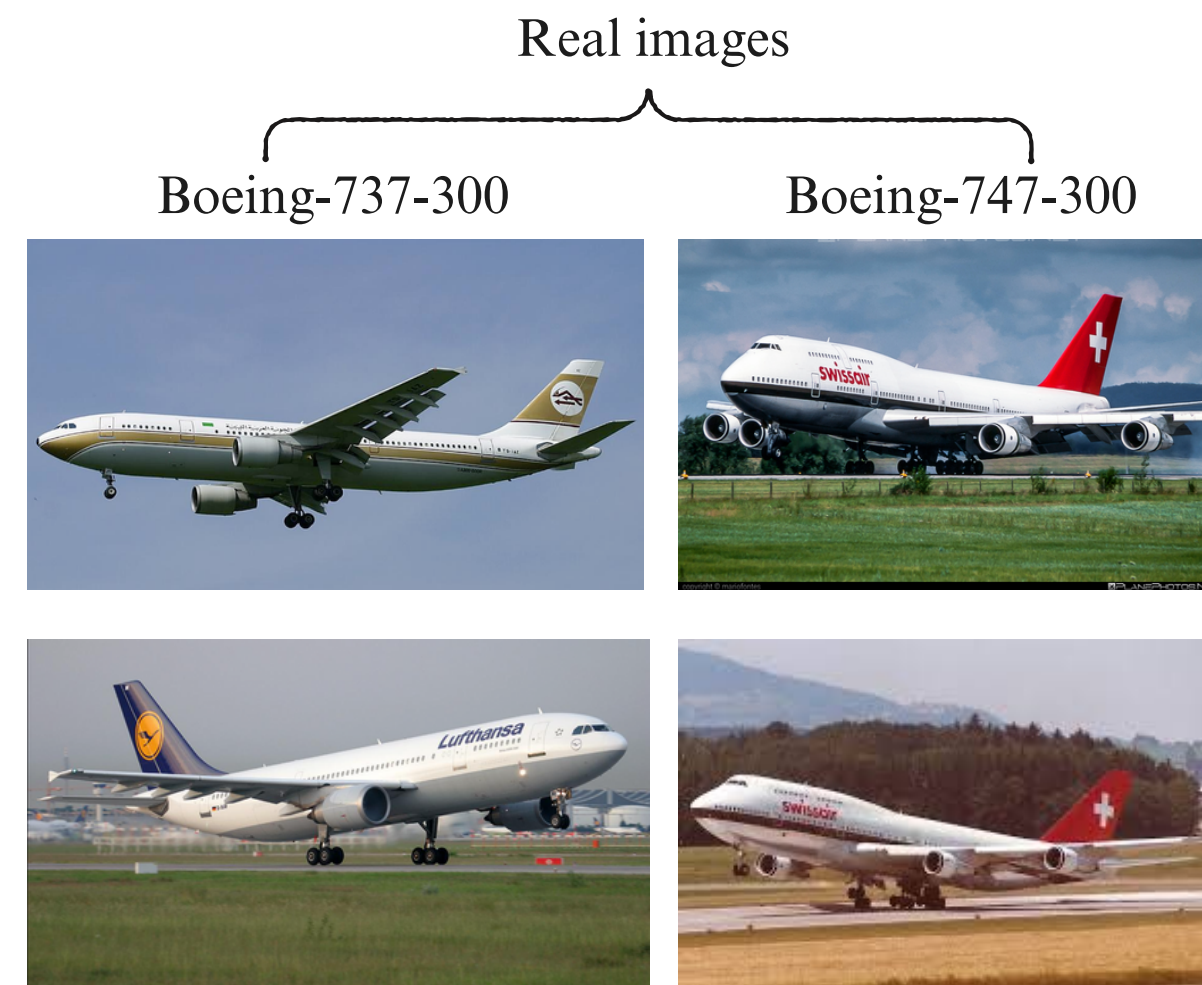**Real** Boeing-737-300

Real Boeing-747-300

Why is FGVC more difficult?

- **Differences between classes** are usually **minor**, yet __important__

- Text-to-image **diffusion models** (for now) usually **don't generate them very well**

Real images

Boeing-737-300

Boeing-747-300

"An airplane of type Boeing-737-300"

Why is FGVC more difficult?

So, just use Img2Img...?

Why is FGVC more difficult?

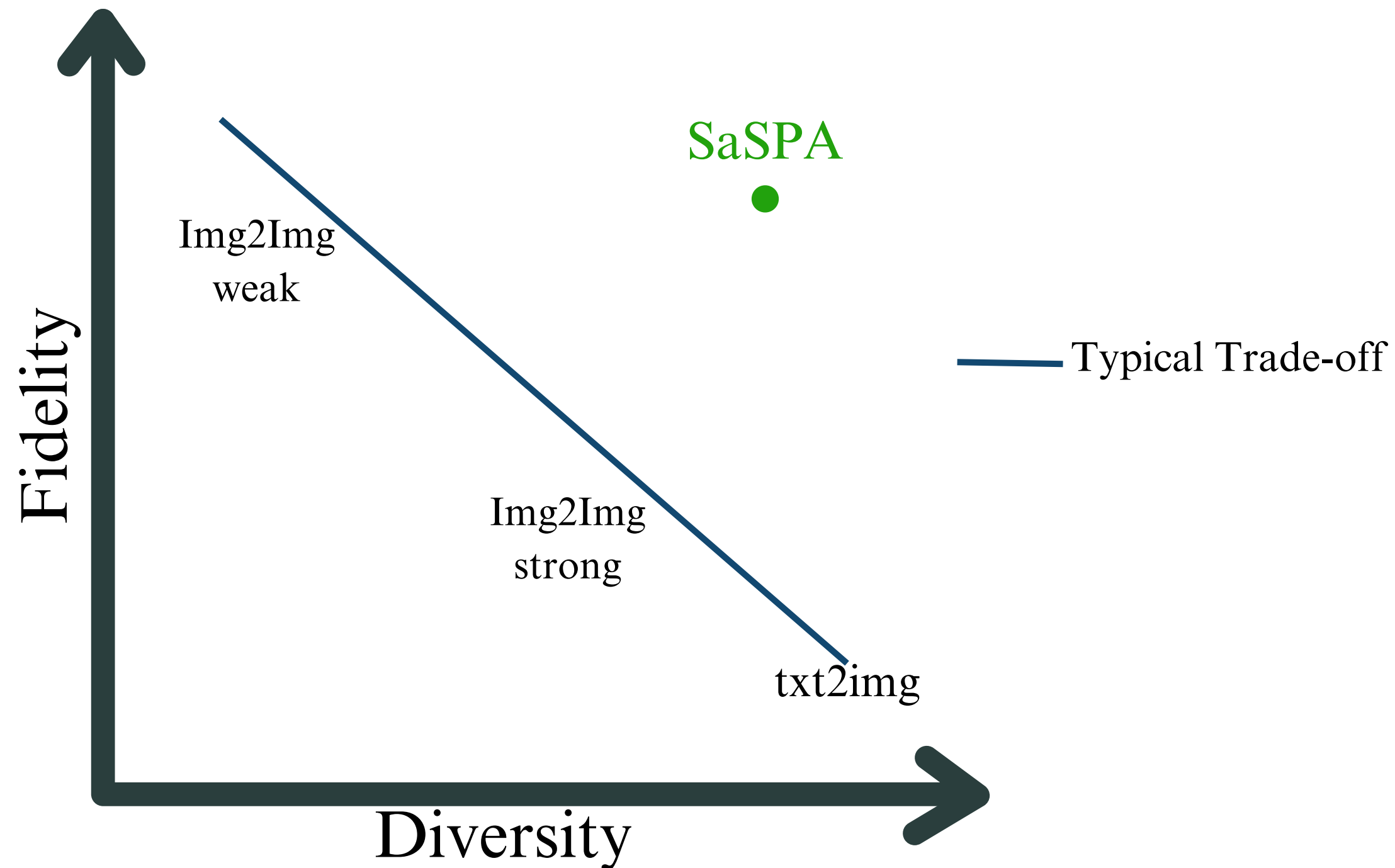So, just use Img2Img...?

Original

Img2Img (s = 0.5)

Img2Img (s = 0.75)

Why is FGVC more difficult?

So, just use Img2Img...?

Results in a trade off:

How to improve?

We looked for an augmentation method, that will:

- Will (1) introduce **significant diversity** while (2) **correctly representing** the **fine-grained class**

- Will not include fine-tuning of the generation model to avoid the heavy computational and time costs

💡 Insight 1:
Using Edge Guidance

- Using **Edge Guidance** as a **prior condition** to **keep the structure, <u>instead of image</u>**

- Advantages:

  - **Better representation**

  - **More DoF** for the model (more **diversity**)
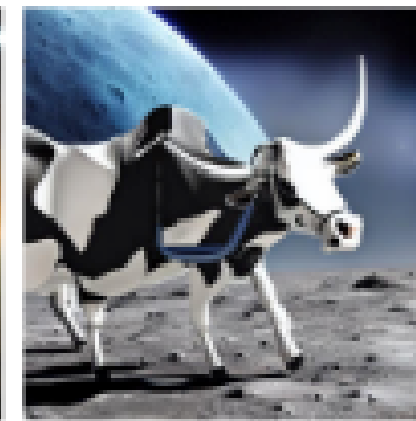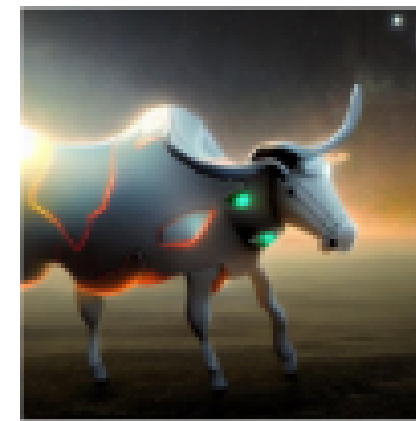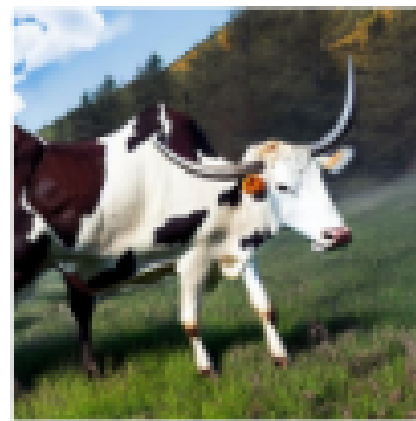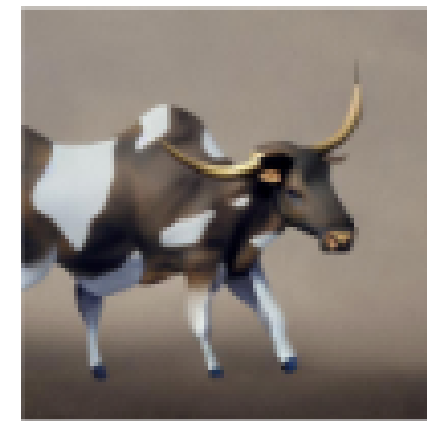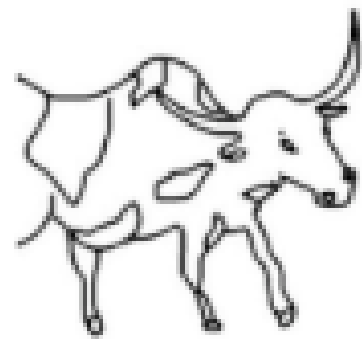
💡Insight 2: Condition on Subject Representation

- Use **subject representation** to **correctly represent other features too**

- For most datasets, **using edge guidance** and **subject representation** allows us to **free the model** from **adhering to a specific image**

- This **significantly** increases **diversity**!

Zero-Shot Subject-driven Generation

Input Image

dog

in batman suit

happily jumping

painting by Van Gogh

in a bucket

swimming underwater

eating ramen

SaSPA
Visual Examples

Original   Augmentations

Cars

CUB

DTD

Results

Type	Augmentation Method	Aircraft	CompCars	Cars	CUB	DTD
	No Aug	81.4	67.0	91.8	81.5	68.5
	CAL-Aug	<u>84.9</u>	70.5	92.4	<u>82.5</u>	<u>69.7</u>
Traditional	RandAug	83.7	72.5	92.6	81.5	69.3
	CutMix	81.8	66.9	91.7	81.8	69.2
	CAL-Aug + CutMix	84.5	70.2	<u>92.7</u>	82.4	<u>69.7</u>
	RandAug + CutMix	84.0	<u>72.6</u>	<u>92.7</u>	81.2	69.2

Results

Real-Guidance (ICLR 2023): Low strength Img2Img with GLIDE

ALIA (NeurIPS 2023): GPT(image captions) + specific strength Img2Img (SDEdit or InstructPix2Pix)

Type	Augmentation Method	Aircraft	CompCars	Cars	CUB	DTD
Traditional	No Aug	81.4	67.0	91.8	81.5	68.5
	CAL-Aug	84.9	70.5	92.4	82.5	69.7
	RandAug	83.7	72.5	92.6	81.5	69.3
	CutMix	81.8	66.9	91.7	81.8	69.2
	CAL-Aug + CutMix	84.5	70.2	92.7	82.4	69.7
	RandAug + CutMix	84.0	72.6	92.7	81.2	69.2
Generative	Real Guidance	84.8	73.1	92.9	82.8	68.5
	ALIA	83.1	72.9	92.6	82.0	69.1

Results

<u>Real-Guidance</u> (ICLR 2023): Low strength Img2Img with GLIDE

<u>ALIA</u> (NeurIPS 2023): GPT(image captions) + specific strength Img2Img (SDEdit or InstructPix2Pix)

Type	Augmentation Method	Aircraft	CompCars	Cars	CUB	DTD
Traditional	No Aug	81.4	67.0	91.8	81.5	68.5
	CAL-Aug	84.9	70.5	92.4	82.5	69.7
	RandAug	83.7	72.5	92.6	81.5	69.3
	CutMix	81.8	66.9	91.7	81.8	69.2
	CAL-Aug + CutMix	84.5	70.2	92.7	82.4	69.7
	RandAug + CutMix	84.0	72.6	92.7	81.2	69.2
Generative	Real Guidance	84.8	73.1	92.9	82.8	68.5
	ALIA	83.1	72.9	92.6	82.0	69.1
Ours	SaSPA w/o BLIP-diffusion	**87.4**	74.8	93.7	83.0	69.8
	SaSPA	86.6	**76.2**	**93.8**	**83.2**	**71.9**

End

Thanks!

Computer Science MLDS School

Eyal Michaeli, Ohad Fried