Effective Data Augmentation With Diffusion Models

Brandon Trabucco [1], Kyle Doherty [2], Max Gurinas [3], Ruslan Salakhutdinov [1]

[1] CMU , [2] MPG Ranch , [3] University of Chicago Laboratory Schools

Machine Learning Department | Carnegie Mellon University School of Computer Science

Data Augmentation Is An Effective Tool

Horizontal Flip

RandAugment
(Cubuk et al., 2019)

CutMix
(Yun et al., 2019)

[1] Cubuk et al., RandAugment: Practical automated data augmentation with a reduced search space, NeurIPS 2020.
[2] Yun et al., CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features, ICCV 2020.

Data Augmentation Is An Effective Tool

Horizontal Flip

RandAugment
(Cubuk et al., 2019)

CutMix
(Yun et al., 2019)

- But, augmentations currently **require a good intuition** about your dataset.

[1] Cubuk et al., RandAugment: Practical automated data augmentation with a reduced search space, NeurIPS 2020.
[2] Yun et al., CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features, ICCV 2020.

We Need Augmentations That **Adapt To Your Dataset**

[1] Cubuk et al., RandAugment: Practical automated data augmentation with a reduced search space, NeurIPS 2020.
[2] Yun et al., CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features, ICCV 2020.

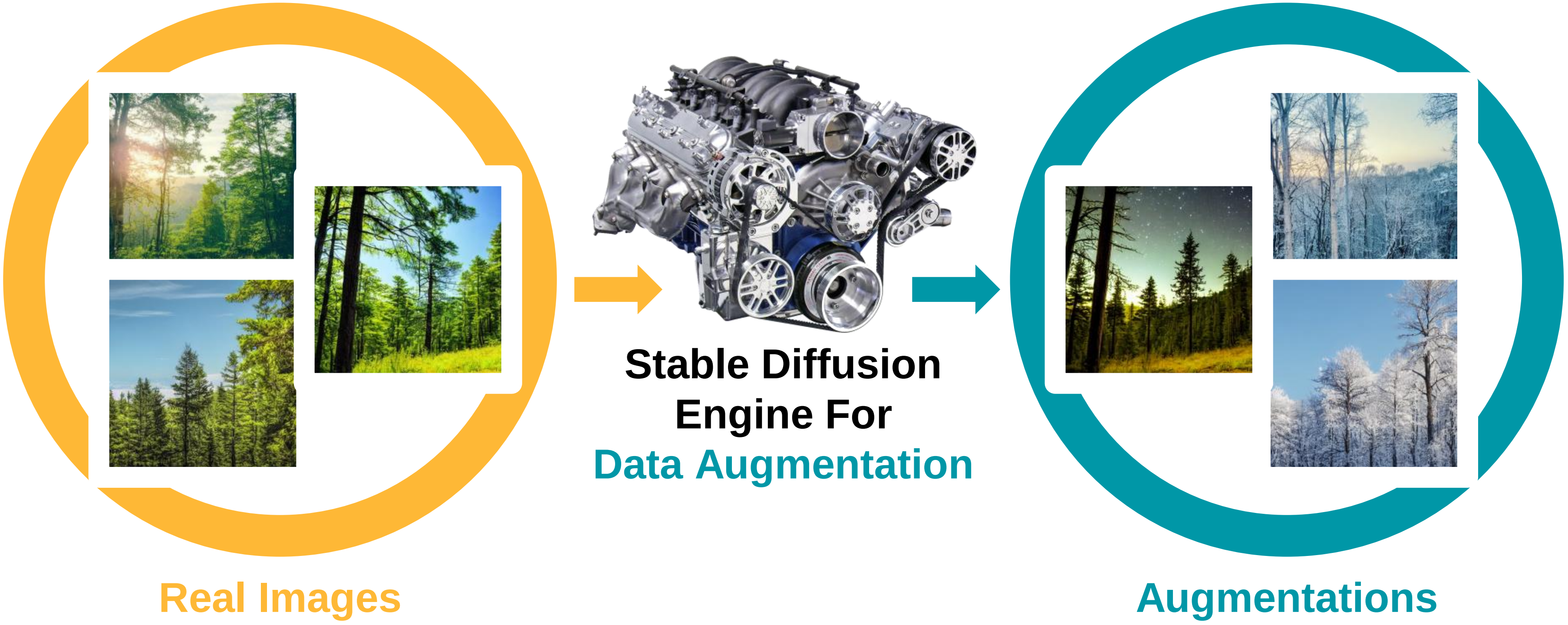
We Need Augmentations That **Adapt To Your Dataset**

[1] Cubuk et al., RandAugment: Practical automated data augmentation with a reduced search space, NeurIPS 2020.
[2] Yun et al., CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features, ICCV 2020.

Real Images

Real Images

Augmentations


Stable Diffusion
Engine For
Data Augmentation
Real Images
Augmentations

DA-Fusion: Data Augmentation With Diffusion

- **Key idea**: **shared context** in your images controls the augmentation.

[3] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.
[7] Rinon, Gal, et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, CVPR 2022.

DA-Fusion: Data Augmentation With Diffusion

- **Key idea**: **shared context** in your images controls the augmentation.

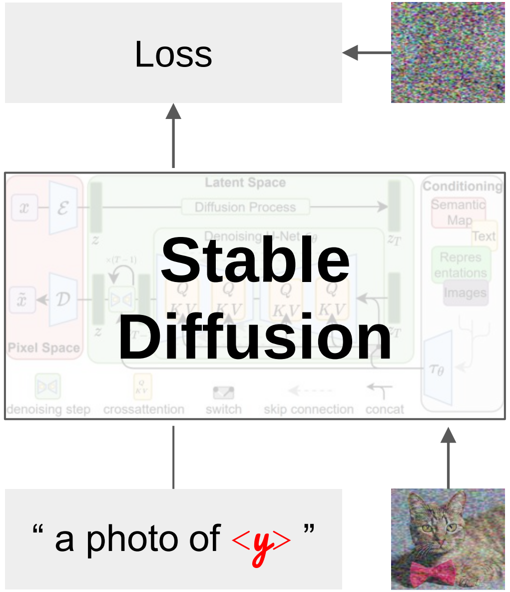
$\max_{y} \log P($  $\mid \text{a photo of a } y\ )$

[3] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.
[7] Rinon, Gal, et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, CVPR 2022.

DA-Fusion: Data Augmentation With Diffusion

- **Key idea**: **shared context** in your images controls the augmentation.

$$\max_{y} \log P($$  $$| \text{ a photo of a } y \,)$$

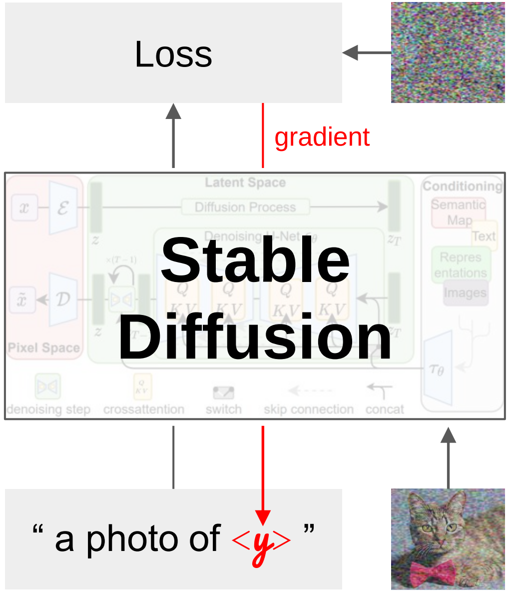


[3] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.
[7] Rinon, Gal, et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, CVPR 2022.

DA-Fusion: Data Augmentation With Diffusion

- **Key idea**: **shared context** in your images controls the augmentation.

$\max_{y} \log P($  $|$ a photo of a y $)$



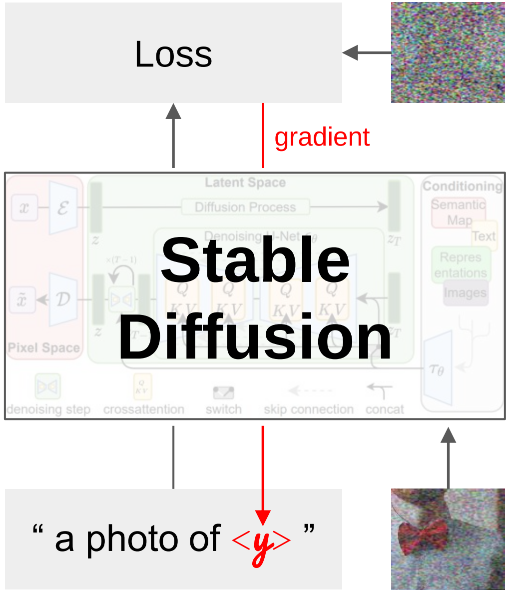
<y> ~ cat wearing a red bow-tie

[3] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.
[7] Rinon, Gal, et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, CVPR 2022.

DA-Fusion: Data Augmentation With Diffusion

- **Key idea**: **shared context** in your images controls the augmentation.

$\max_{y} \log P($  $\mid \text{a photo of a } y \;)$



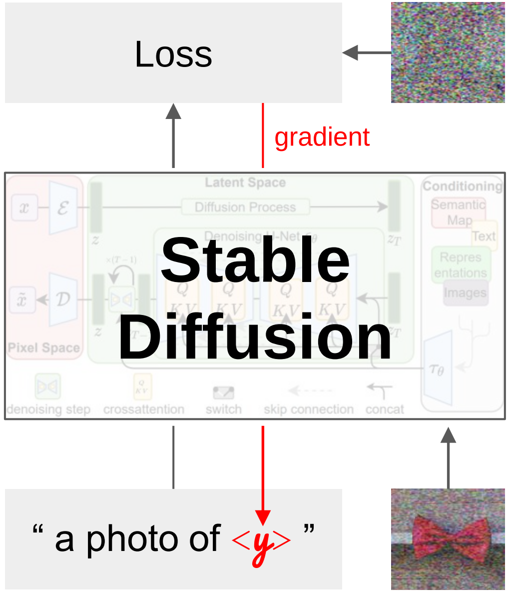
<y> ~ wearing a red bow-tie

[3] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.
[7] Rinon, Gal, et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, CVPR 2022.

DA-Fusion: Data Augmentation With Diffusion

- **Key idea**: **shared context** in your images controls the augmentation.

$\max_{y} \log P($  $\mid \text{a photo of a } y \,)$



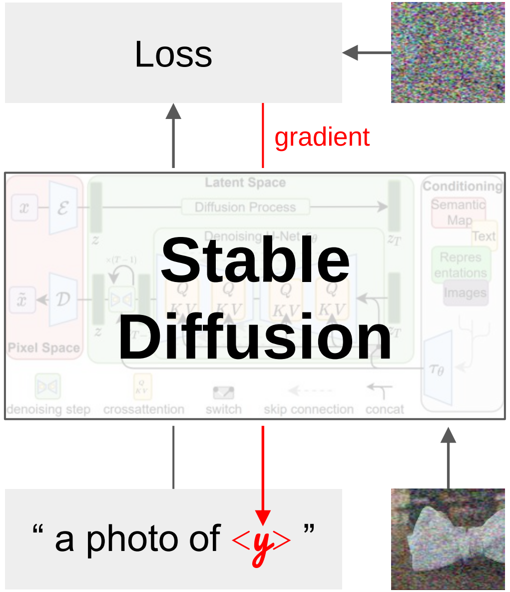
<y> ~ red bow-tie

[3] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.
[7] Rinon, Gal, et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, CVPR 2022.

DA-Fusion: Data Augmentation With Diffusion

- **Key idea**: **shared context** in your images controls the augmentation.

$\max_{y} \log P($  $\mid$ a photo of a y $)$



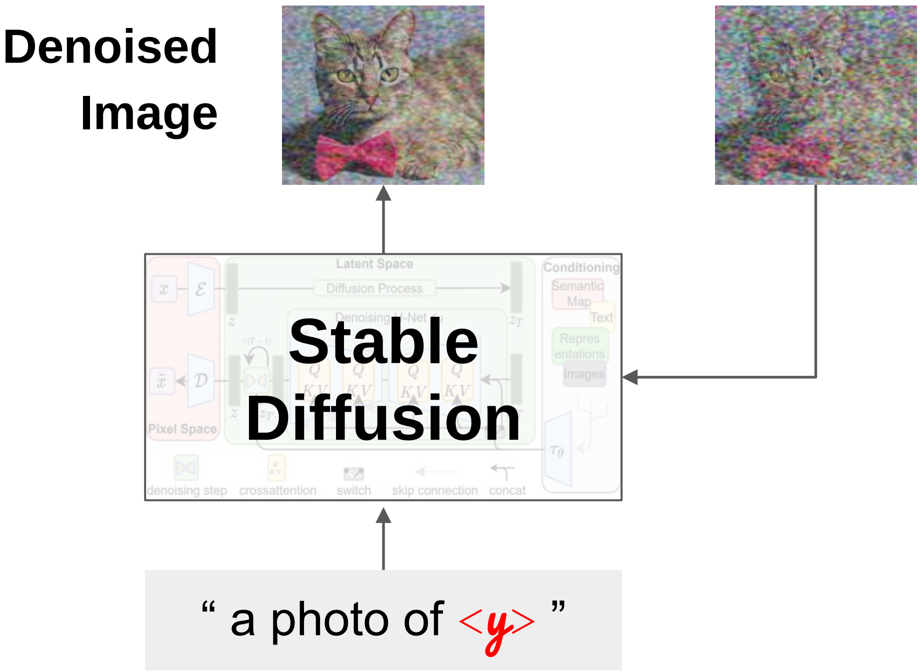
<y> ~ bow-tie

[3] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.
[7] Rinon, Gal, et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, CVPR 2022.

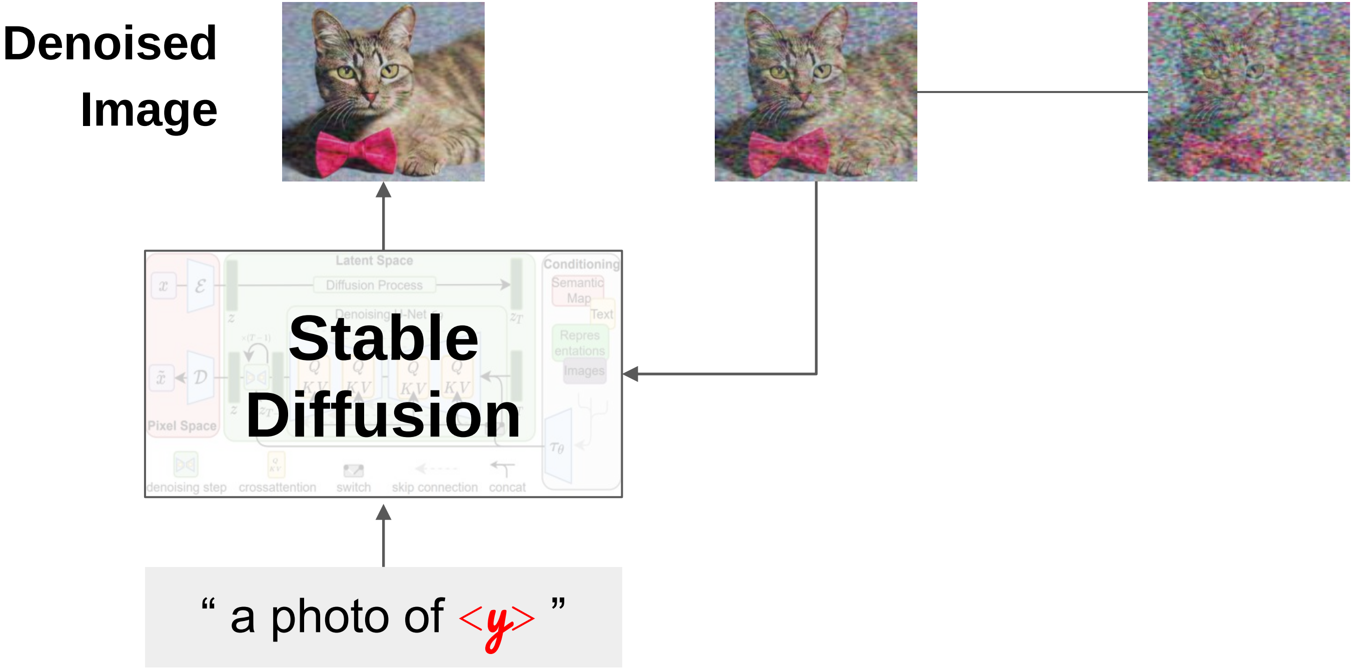
DA-Fusion: Data Augmentation With Diffusion

Source Image

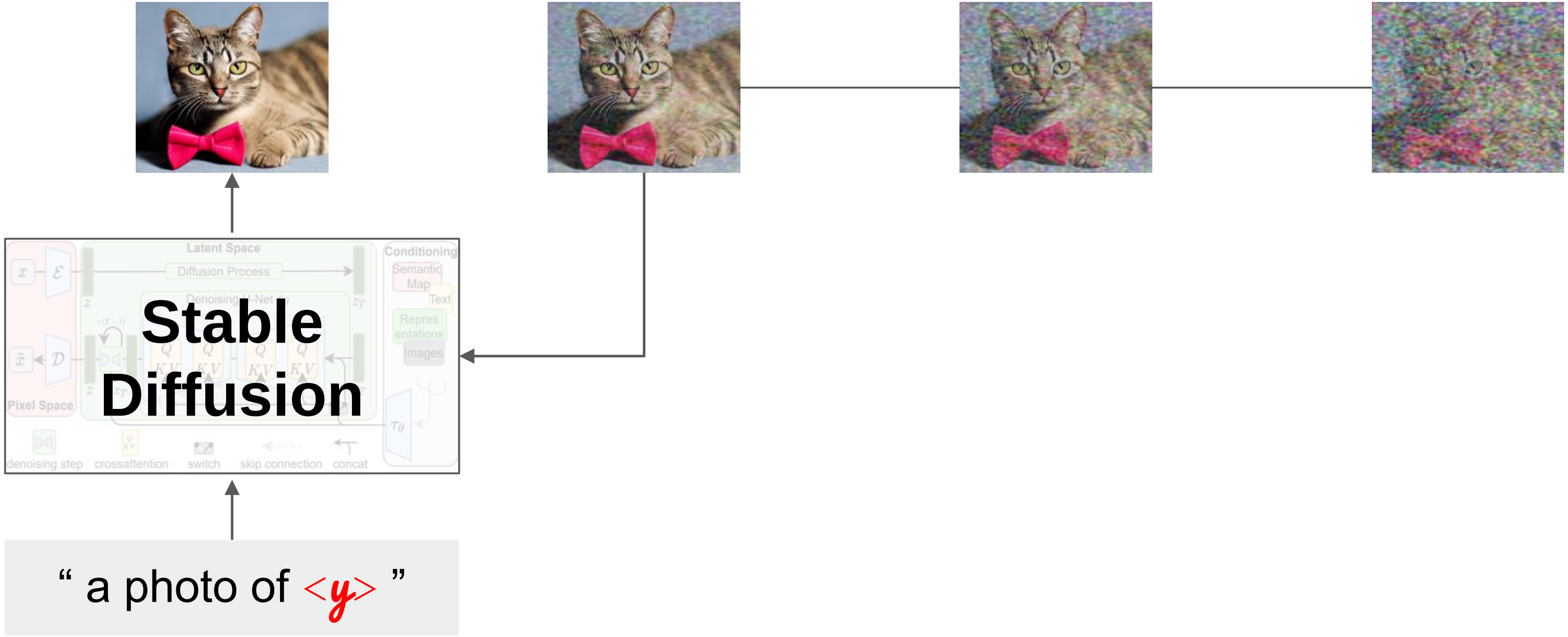
DA-Fusion: Data Augmentation With Diffusion



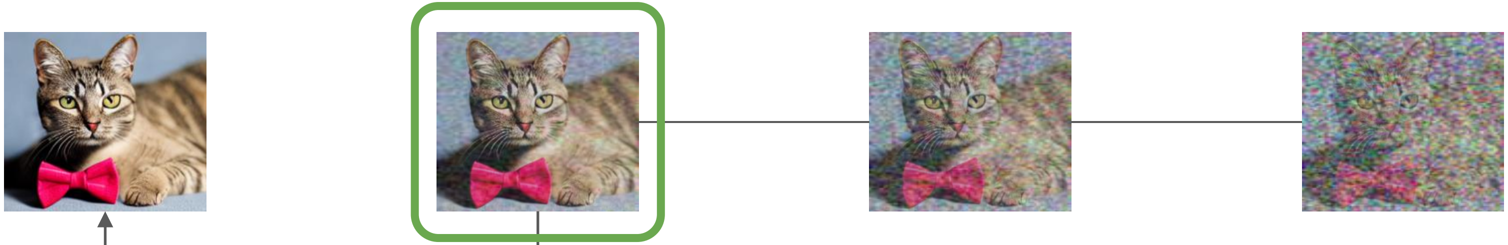
DA-Fusion: Data Augmentation With Diffusion

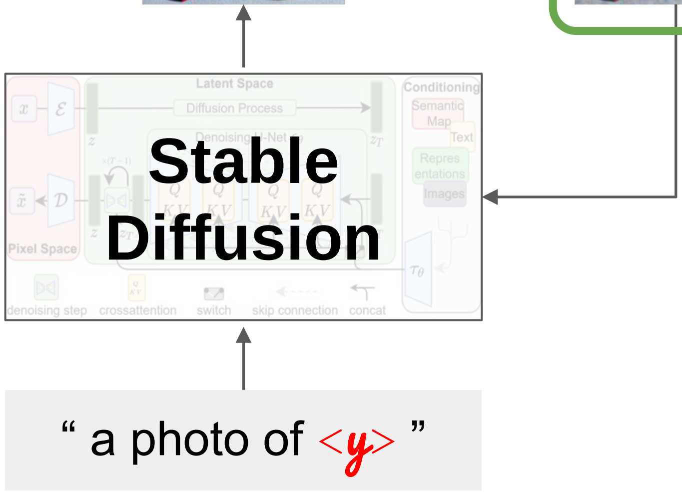


DA-Fusion: Data Augmentation With Diffusion



DA-Fusion: Data Augmentation With Diffusion

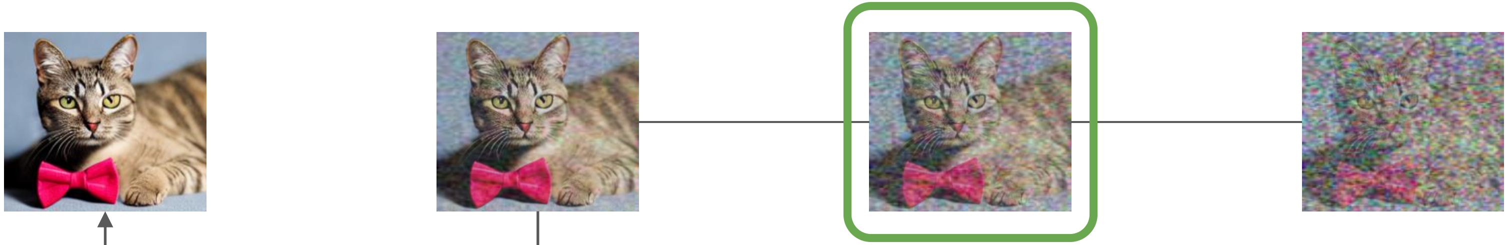


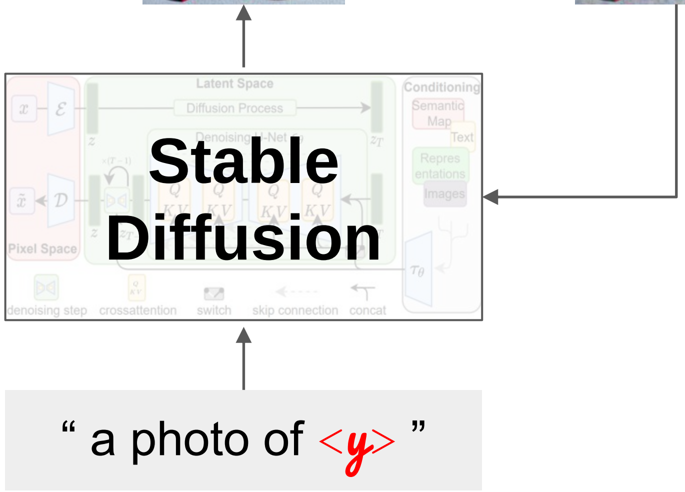


- Control the augmentation extent with the **insertion timestep**

Low

DA-Fusion: Data Augmentation With Diffusion

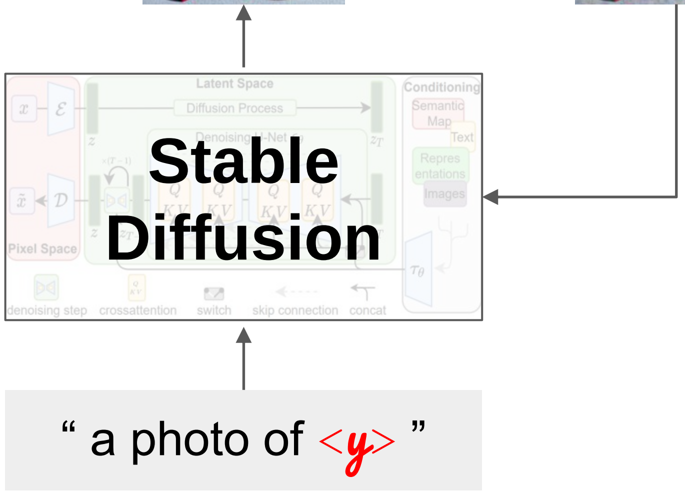




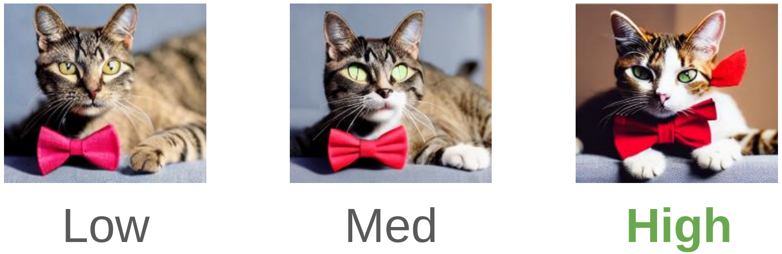
- Control the augmentation extent with the **insertion timestep**

Low **Med**

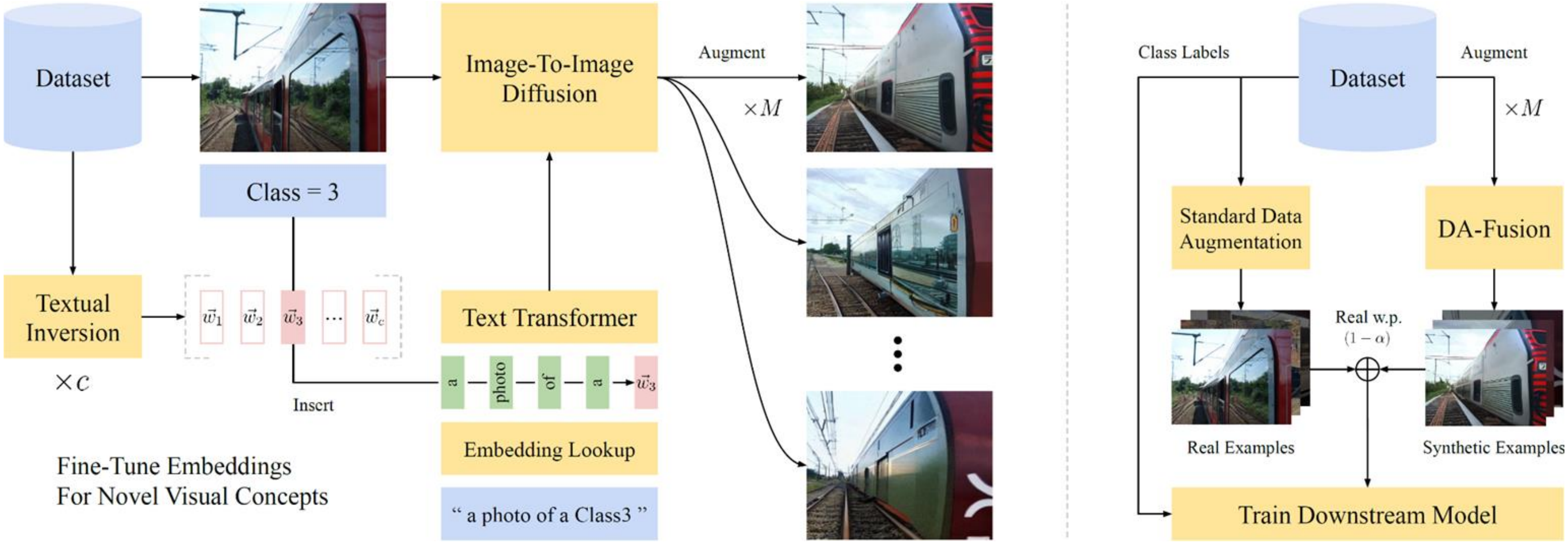
DA-Fusion: Data Augmentation With Diffusion



- Control the augmentation extent with the **insertion timestep**



DA-Fusion: Data Augmentation With Diffusion



[7] Rinon, Gal, et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, CVPR 2022.
[8] Chenlin, Meng, et al., SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations, ICLR 2022.






How Do We Evaluate DA-Fusion?

Question: How much do augmentations from DA-Fusion improve classification?

How Do We Evaluate DA-Fusion?

Question: How much do augmentations from DA-Fusion improve classification?

- Six few-shot classification tasks from literature

Common Concepts

Fine-Grain Concepts

How Do We Evaluate DA-Fusion?

Question: How much do augmentations from DA-Fusion improve classification?

- Six few-shot classification tasks from literature and **one we contribute**.

Common Concepts

Fine-Grain Concepts

Leafy Spurge

Novel Concepts

How Do We Evaluate DA-Fusion?

Question: How much do augmentations from DA-Fusion improve classification?

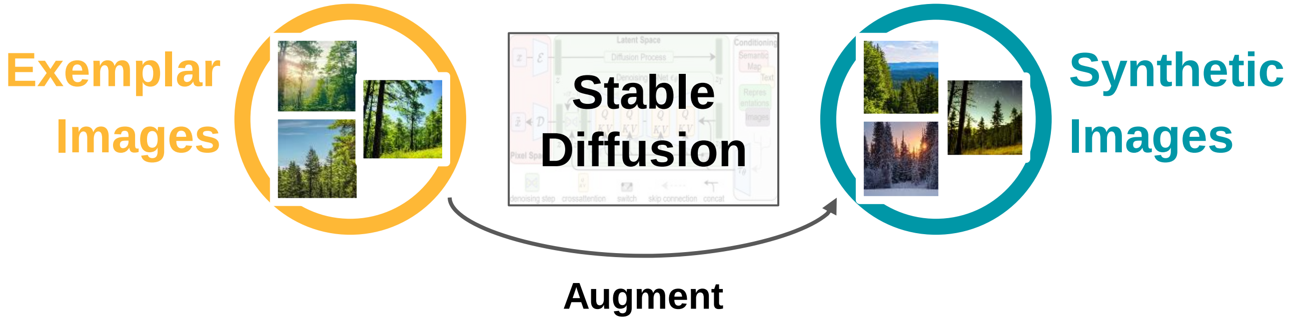
- Given a handful of **real images**

Exemplar Images

How Do We Evaluate DA-Fusion?

Question: How much do augmentations from DA-Fusion improve classification?

- Given a handful of **real images**, generate **augmentations**

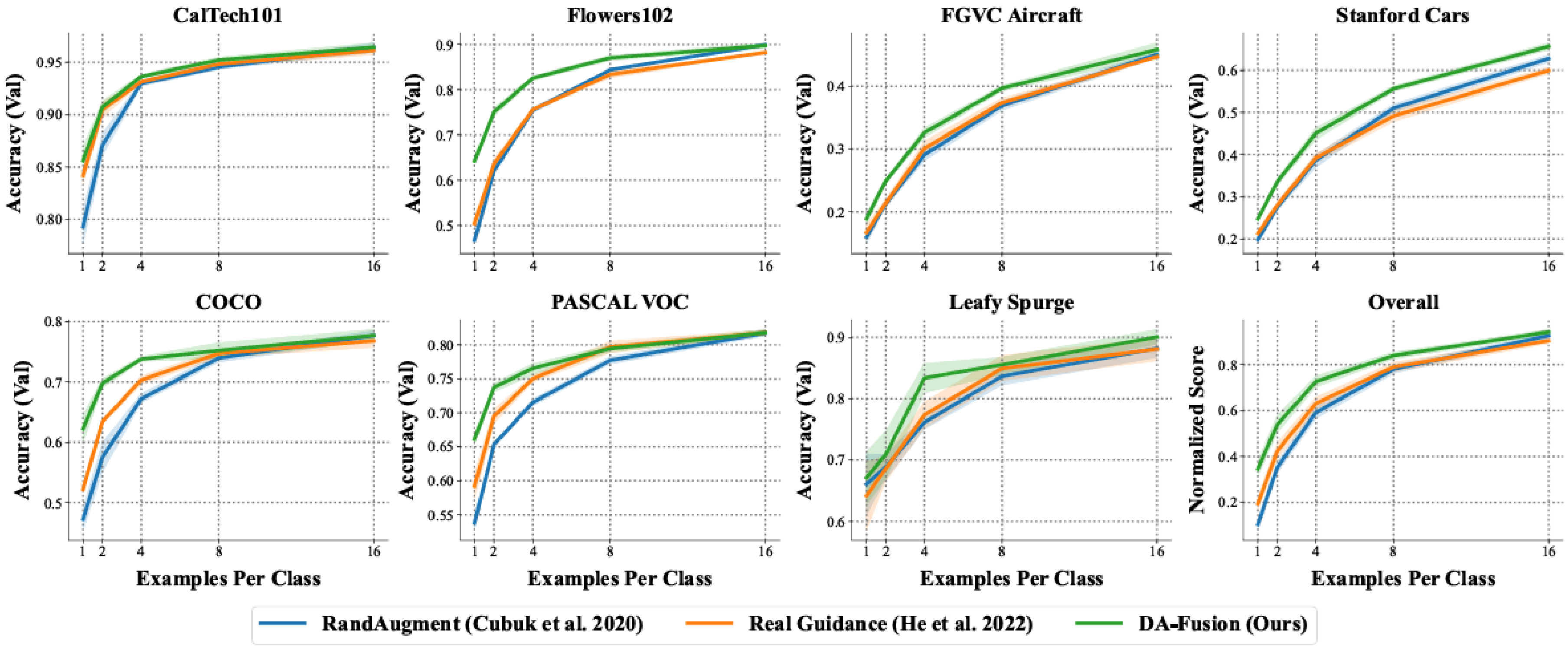


How Do We Evaluate DA-Fusion?

Question: How much do augmentations from DA-Fusion improve classification?

- Given a handful of **real images**, generate **augmentations**
- Train classifiers on a mix of real and synthetic data

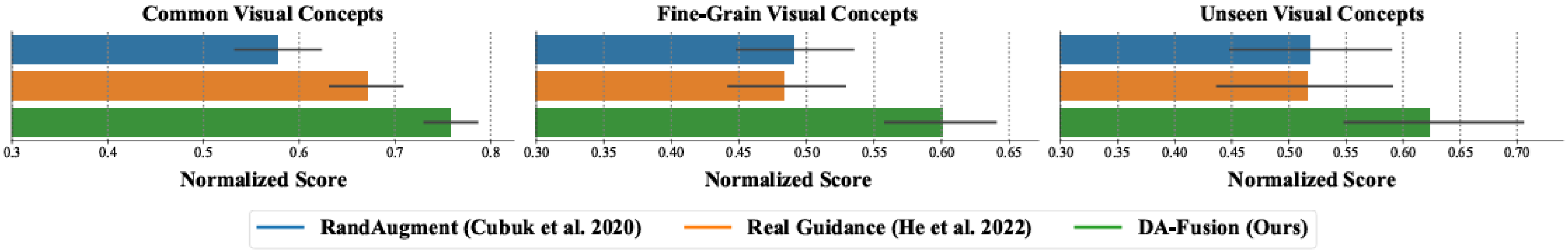
DA-Fusion Improves **Few-Shot Learning**



[1] Cubuk et al., RandAugment: Practical automated data augmentation with a reduced search space, NeurIPS 2020.
[2] He et al., Is synthetic data from generative models ready for image recognition?, ICLR 2023.

DA-Fusion Has **Consistent Performance**

Leafy Spurge



- DA-Fusion has **strong performance** for **all types of concepts**.

[1] Cubuk et al., RandAugment: Practical automated data augmentation with a reduced search space, NeurIPS 2020.
[2] He et al., Is synthetic data from generative models ready for image recognition?, ICLR 2023.

Strong Performance ✓

Strong Performance ✓

How Do You Control The Augmentation?

When Is Additional Control Necessary?

Real images are often cluttered with **distracting concepts**.

$\max_y \log P($  $\mid \text{a photo of a } y \ )$

When Is Additional Control Necessary?

Real images are often cluttered with **distracting concepts**.

$$\max_y \; \log P($$

$$\mid \text{a photo of a } y \;)$$

Which concept should DA-Fusion generate: cats and/or dogs?

How Do You Control What Concept Is Learned?

Implicit Solution: **select better images** without **distracting concepts**.

$$\max_y \log P($$

$$\mid \text{a photo of a } y\ )$$

How Do You Control What Concept Is Learned?

~~**Implicit Solution**: **select better images** without **distracting concepts**.~~

$$\max_{y} \log P\Big($$

$$\Big| \text{ a photo of a } y \Big)$$

This might be **costly**, what else can we do?

How Do You Control What Concept Is Learned?

Explicit Solution: **prompt with context** about the objects you want ignored.

How Do You Control What Concept Is Learned?

Explicit Solution: **prompt with context** about the objects you want ignored.

- **Why?** Prompts can supplement information **when images have ambiguity**.

How Do You Control What Concept Is Learned?

Explicit Solution: **prompt with context** about the objects you want ignored.

- **Why?** Prompts can supplement information **when images have ambiguity**.

DA-Fusion

How Do You Control What Concept Is Learned?

Explicit Solution: **prompt with context** about the objects you want ignored.

- **Why?** Prompts can supplement information **when images have ambiguity**.

DA-Fusion

\+ “ a <*y*> to the right of the dog ”

DA-Fusion

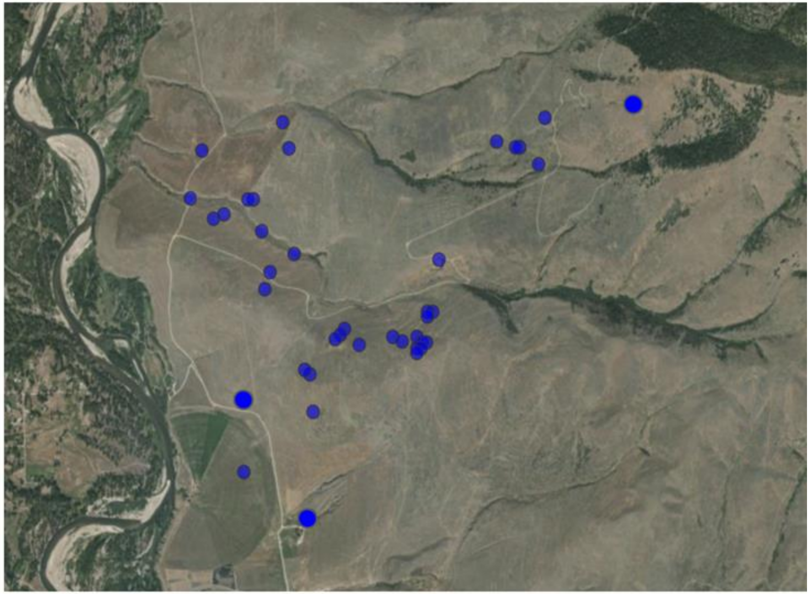
Real-World Evaluation

Leafy spurge (Euphorbia esula): A Problematic Weed in N. America

Photo credit: Montana State University

Locations of spurge surveys in western Montana

- We surveyed 40 sites varied in land-use history and plant community composition

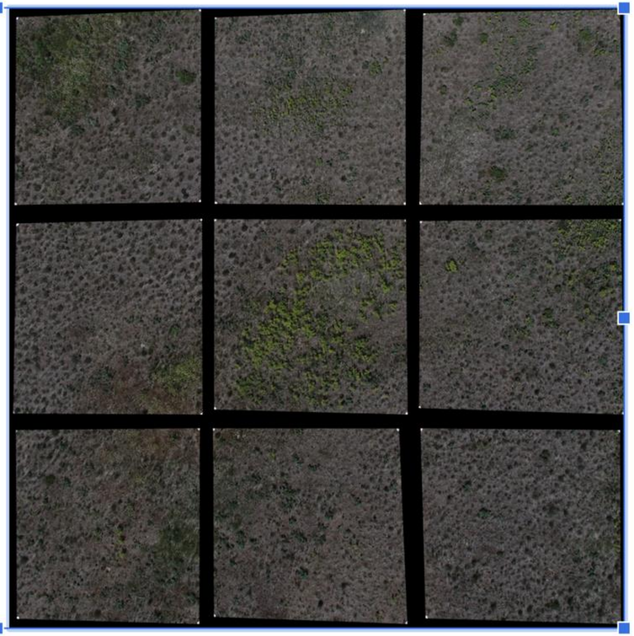


Botanists verified spurge presence at survey sites

- We searched for spurge within nine 10 x 10 meter plots


Post-processing: Identify plot boundaries

- Example: surveyed nine plots at each site. Markers visible in the image were used to crop plots.



Drone then imaged surveyed areas

Post-processing: Four-crop and verify

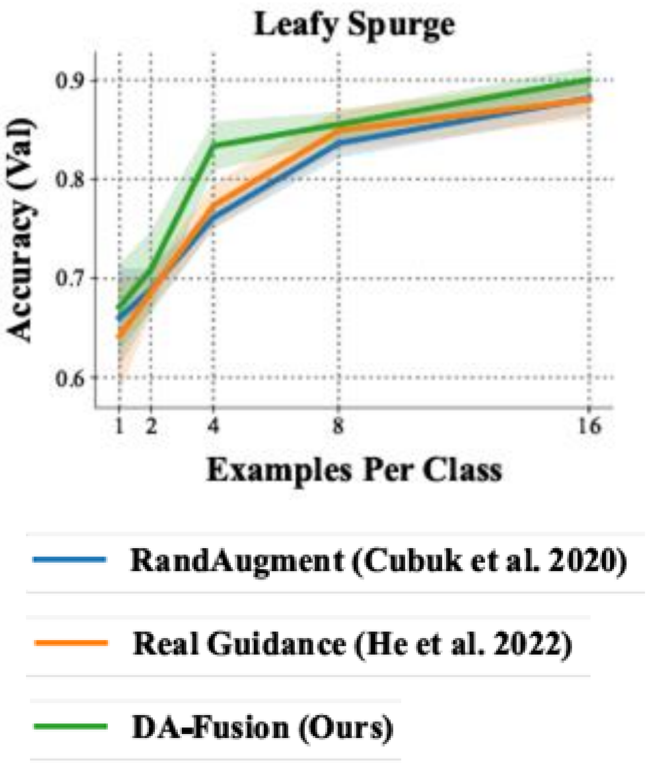
- We further subdivided imagery into quarters of 250 x 250 pixels in size (approximately 3.5 x 3.5 m).

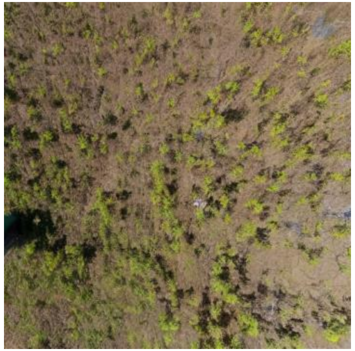
Aerial spurge images distinct from LAION dataset

- Top-down imagery of prairie ecosystems from 50m above the ground.
- Existing outside the domain of LAION and other foundation model training sets. ‘
- This makes these data well-suited for few-shot research.

LAION-5B Schuhmann et al. [2022]

Results + Synthetic Generations





original generation

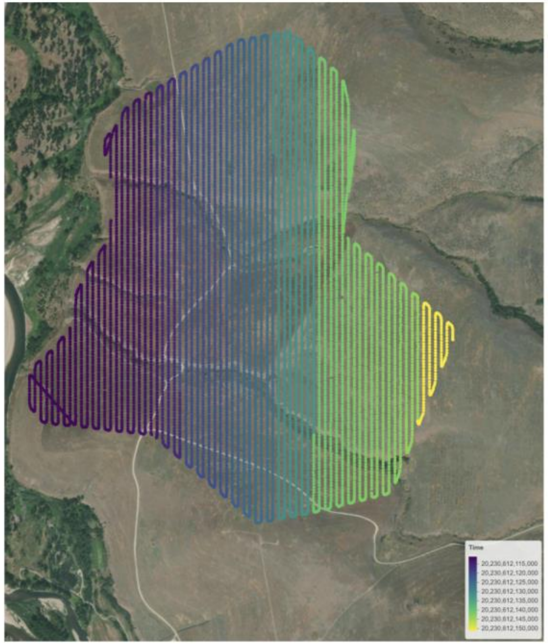
+ mountain

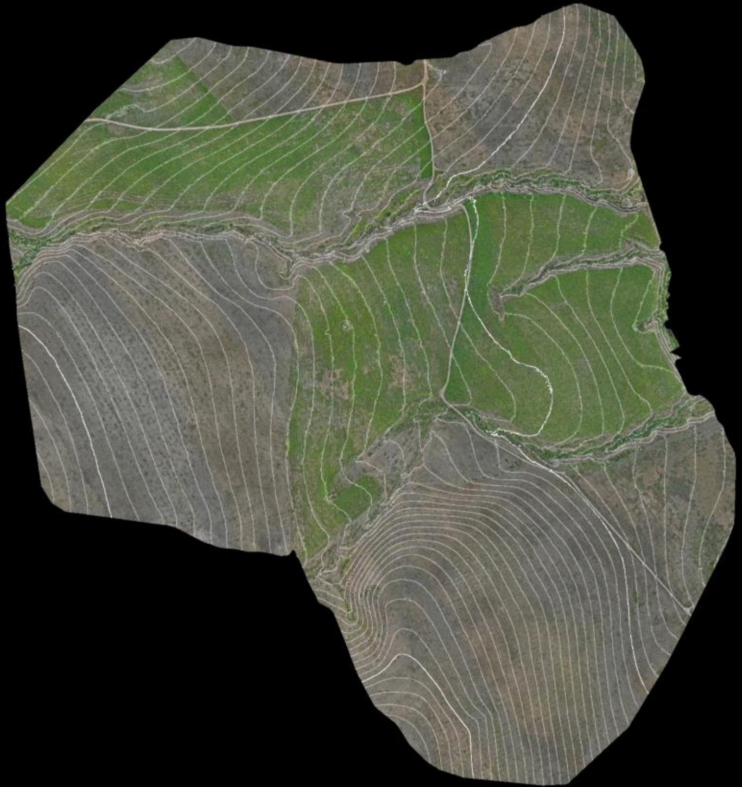
+ grassland

+ snow

Leafy Spurge Dataset V2

- We are working on the next version of the spurge dataset that is well-suited for mapping spurge presence at scale with a public release.



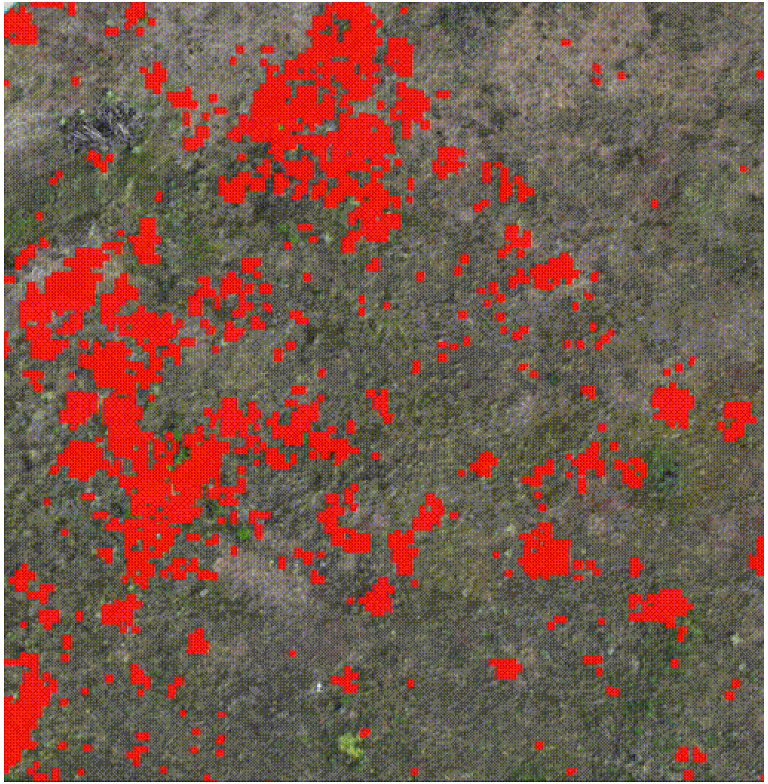


DroneDeploy

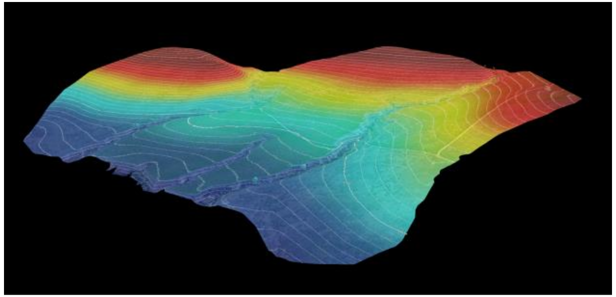
DroneDeploy

Fine-scale Classification Maps

- Fine-scale classification will enable more effective management of leafy spurge and rapidly respond to these outbreaks.
- We hope to extend this approach to other species to monitor ecosystems at scale and rapidly respond to ecological change in near real-time.

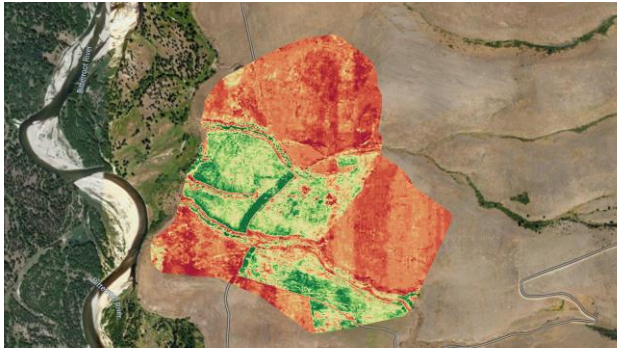


Ecologist Collaborators Offer a Rich Source of Unique Datasets



Elevation paired with RGB

Bear individual retrieval



Non-visible spectra (plant health)

Landscape change detection

Questions


Brandon Trabucco


Kyle Doherty


Max Gurinas


Russ Salakhutdinov

Code

Read more and check out the code: **btrabuc.co/da-fusion**

Thanks For Listening!