Learned Image Augmentation for Distribution Shift Robustness Using GFlowNets

Nathan Kim*, Pratham Soni*, Mo Tiwari

{nathangk, prathams, motiwari}@stanford.edu

*equal contribution

Our Setting

- **Generative Flow Networks** (below, GFlowNets) are a generative network first proposed in [1]. They compose samples using **flows** between **states** and **actions**.
- The **flow property** inherent in GFLowNets mitigate the *probability deserts* observed in MCMC/GAN models.
- Therefore, we claim that GFlowNets are capable of producing **high quality** samples with **enough variance** to meaningfully augment large image datasets.

This work presents the first GFlowNet architecture capable of generating large images and evaluates its ability to augment image datasets.

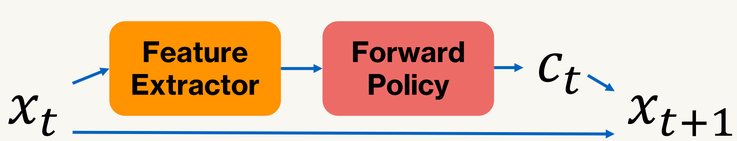
1: GFlowAug

Succinctly, GFlowNets use methodology from **RL** to create generative models with **balanced** sampling distributions.

Objects x are generated by applying a series of **actions** $a_1, \ldots, a_n$ to an initial **state** s_0 to generate a final state $x = s_n$. Together, the actions and states form a **trajectory** τ.

Each action between intermediary states is assigned a **flow** value $F(s \to s')$. If flows are preserved in any trajectory, then the sampling distribution of the objects $P(x)$ is **guaranteed** to be **proportional** to $R(x)$, the unnormalized PDF of the design distribution.

Our image model generates subsets c_t of the full image at each action. The states are thus a concatenation of the **partially generated image** with a mask of observed pixels. The actions are modeled as such:



2: Data Poverty & GFlowNets

Dataset	Pre-train Data Fraction
iWildCam	0.00679
PovertyMap	0.0727
GlobalWheat	0.208

- Optimization steps over **trajectories**, not **images.**
- The number of **trajectories** which can be sampled in training is exponential in the number of training images.
- **Therefore,** augmenters can be trained with a fraction (up to 150x reduction) of the original training data.

3: Augmentation Performance

Dataset	Leaderboard (std.)	Baseline	GFlowAug
iWildCam	32.8 (0.1)	32.79	**33.27**
PovertyMap	0.49 (0.06)	0.3561	0.3751
GlobalWheat	51.2 (1.8)	48.87	50.86

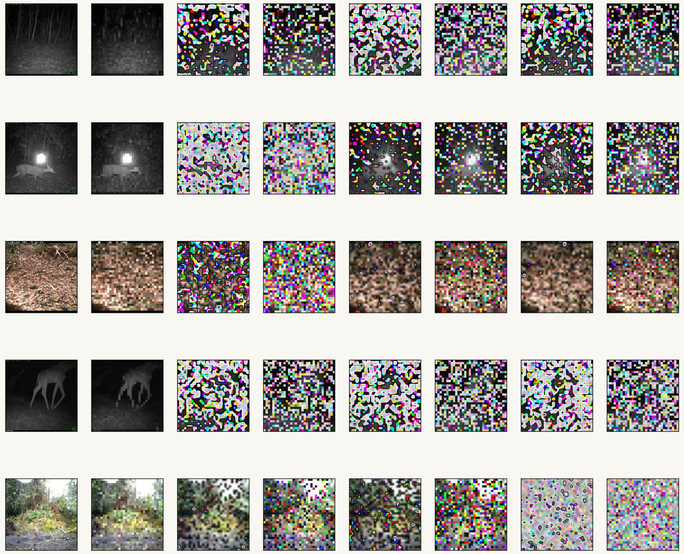
Table 3. Table of results comparing GFlowAug results against both published leaderboard values and re-run baselines.

- We evaluate our method by measuring its improvement on baseline metrics across three image tasks in the WILDS [2] benchmark: **iWildCam** (metric: avg. F1), **PovertyMap** (metric: Pearson *r*), and **GlobalWheat** (metric: top-1 accuracy).
- Baseline models all improve when augmented with GFlowAug, and iWildCam is improved over the **leaderboard average**.

Key takeaway: **GFlowNets can produce image reconstructions that improve model generalization using a fraction of the original dataset.**

4: Qualitative Sample Evaluation

- Downsampling of input images can place a bottleneck on the reconstruction quality (noise ranges from minimal to near-complete).
- Good reconstruction of **textures** (row 3); struggles in **poor light** (rows 1/2/4)



Future Work + Acknowledgements

1. [3] decodes images into their Haar decompositions before generating them; investigate how this improves the internal consistency of GFlowAug samples.
2. Improve policy network to focus on informationally-dense locations for generation first.
3. Thank you to Mo and the Thrun Lab for all their kindness and support!

[1] Bengio, Emmanuel, et al. "Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation." Advances in Neural Information Processing Systems 34 (2021).
[2] Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." International Conference on Machine Learning. PMLR, 2021.
[3] Yu, Jason J., Konstantinos G. Derpanis, and Marcus A. Brubaker. "Wavelet flow: Fast training of high resolution normalizing flows." Advances in Neural Information Processing Systems 33 (2020): 6184-6196.