Color Transfer with Modulated Flows

M. Larchenko, A. Lobashev, D. Guskov, V. Palyulin

Skolkovo Institute of Science and Technology, Moscow

Abstract

Here we introduce Modulated Flows (ModFlows), a novel approach for color style transfer between images based on rectified flows. The primary goal of the color transfer is to adjust the colors of a target image to match the color distribution of a reference image. Our technique is based on optimal transport and executes color transfer as an invertible transformation within the RGB color space.

Qualitative Comparison with Existing Methods

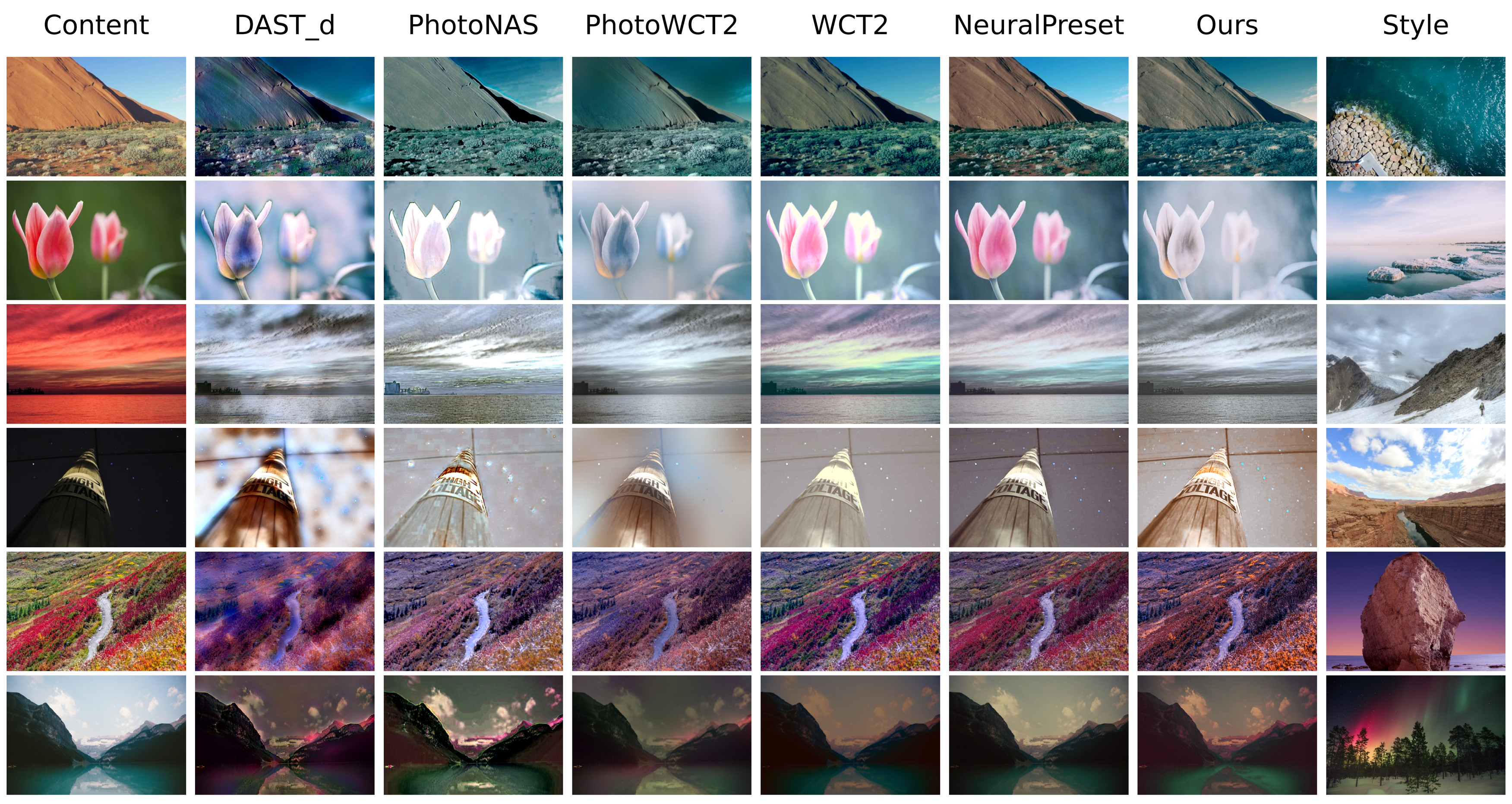
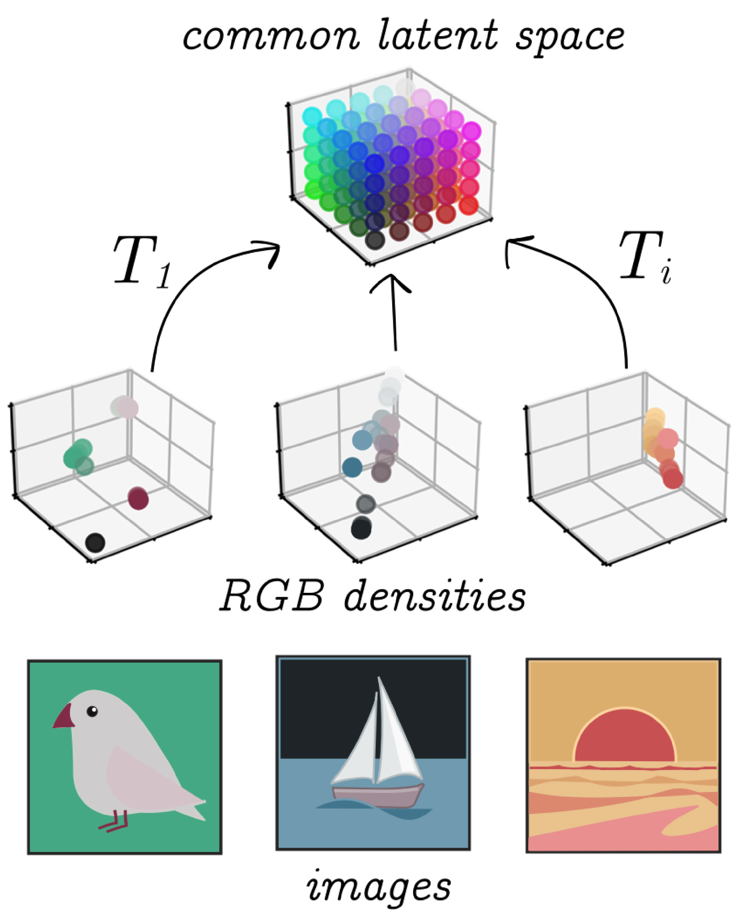


Figure: Examples from Unsplash Lite test set. Our model achieves the state-of-the-art performance in terms of content and style similarity, i.e the most exact match with the reference palette without visible distortion.

General Scheme

Dataset of rectified flows

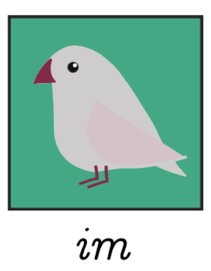


First, we generate a set of 5826 flow-image pairs. Each flow T is trained to map RGB density of an image into the uniform cube.

$$T(x_0) = x_0 + \int_{t=0}^{1} v_{im}\,(x_t\,,\,t)\,dt$$

Note, that a rectified flow T is bijective.

Encoder training

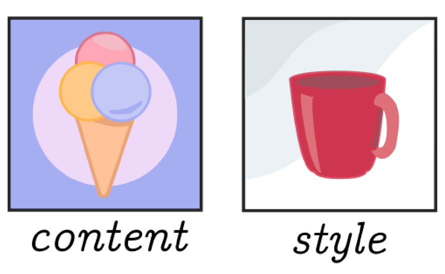


$im \to$ Encoder $\to e$

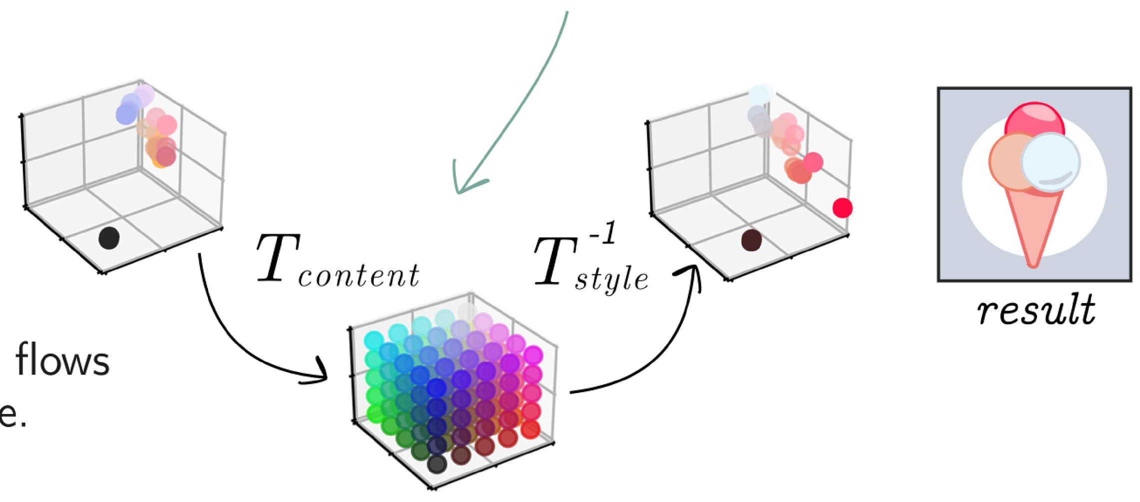
Then we train an encoder model to produce weights e for T

$$loss \sim \mathbf{E}\left[\underbrace{v_e\,(x_t\,,\,t)}_{T \text{ predicted by encoder}} - \underbrace{v_{im}\,(x_t\,,\,t)}_{T_{im} \text{ from the dataset}}\right]^2$$

Inference



$\to$ Encoder $\to e_{content}\,,\, e_{style}$



Color style transfer is a composition of a forward content and an inverse style flows applied to the content image.

Algorithm

1. Produce a dataset of flow-image pairs, where flows' weights θ_i are trained to map a color distribution X_i of an image $\mathcal{I}_i$ into the uniform cube U. We follow *Liu et al. 2022* with an interpolation $X_t = t\,U + (1-t)X_i$

$$\min_{\theta_i} \int_{t=0}^{1} \mathbb{E}_{(U,X_i)\sim\pi_{\text{trivial}}}\left[\|U - X_i - v_{\theta_i}(X_t, t)\|^2\right] dt.$$

2. Train the encoder on batches from the dataset, such that the output vector $\mathrm{Enc}(\mathcal{I}_i) = e_i$ is a flow parametrization for an image $\mathcal{I}_i$.

Algorithm Tuning

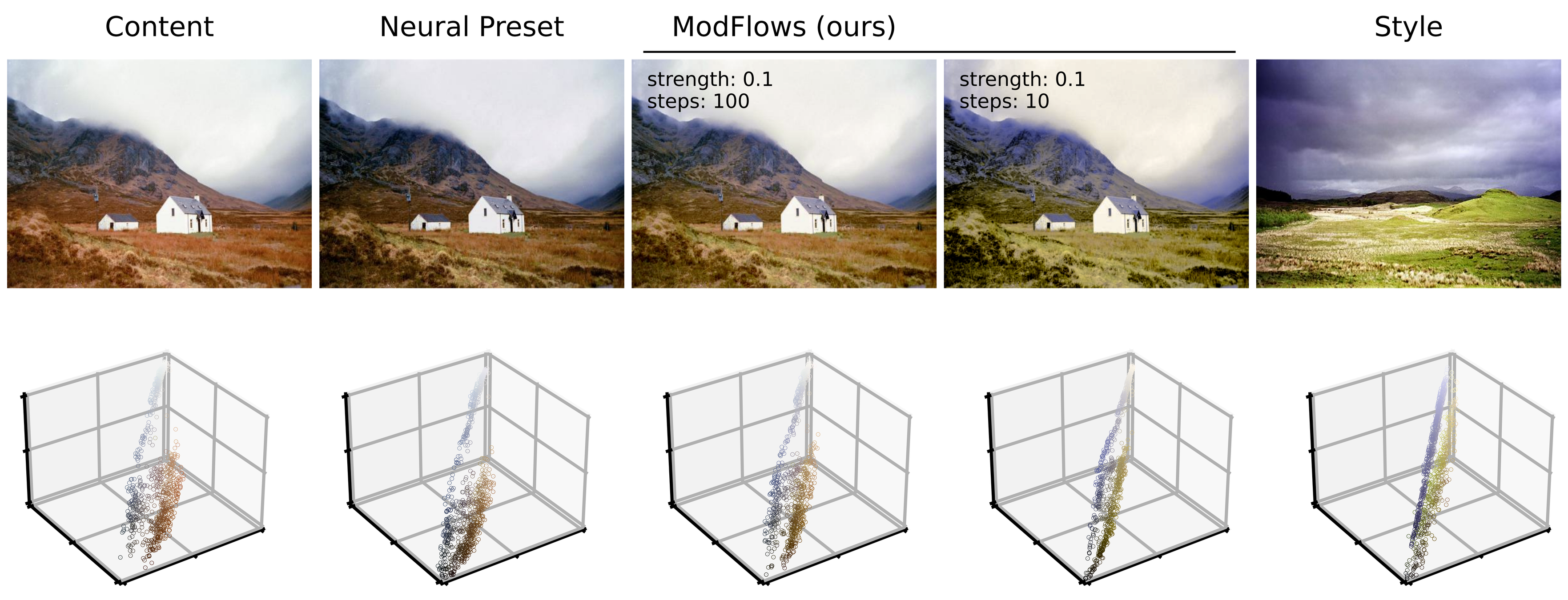


Figure: Variation of a number of steps for ODE solver (steps) and a percent of interpolation curve passed (strength) results in different amount of changes for a distribution.

Comparison Table of Scores

Aggregated scores (DISTS)↓			
Algorithm	Grayscale	Depth	Edge
ModFlows (ours)	**0.129**	**0.217**	0.220
MKL	0.146	0.227	0.224
CT	0.169	0.234	0.232
WCT2	0.170	0.228	0.249
PhotoWCT2	0.191	0.236	**0.217**
DAST_d	0.204	0.267	0.224
DAST_da	0.214	0.282	0.229
PhotoNAS	0.224	0.276	0.270
Deep Preset	0.384	0.400	0.387

Style distance↓	
Algorithm	mean $\pm$ std
DAST_d	**0.112** $\pm$ **0.039**
ModFlows (ours)	0.123 $\pm$ 0.049
DAST_da	0.127 $\pm$ 0.042
PhotoWCT2	0.129 $\pm$ 0.055
MKL	0.145 $\pm$ 0.060
WCT2	0.163 $\pm$ 0.065
CT	0.166 $\pm$ 0.064
PhotoNAS	0.183 $\pm$ 0.069
Deep Preset	0.384 $\pm$ 0.171

Content Metrics Study

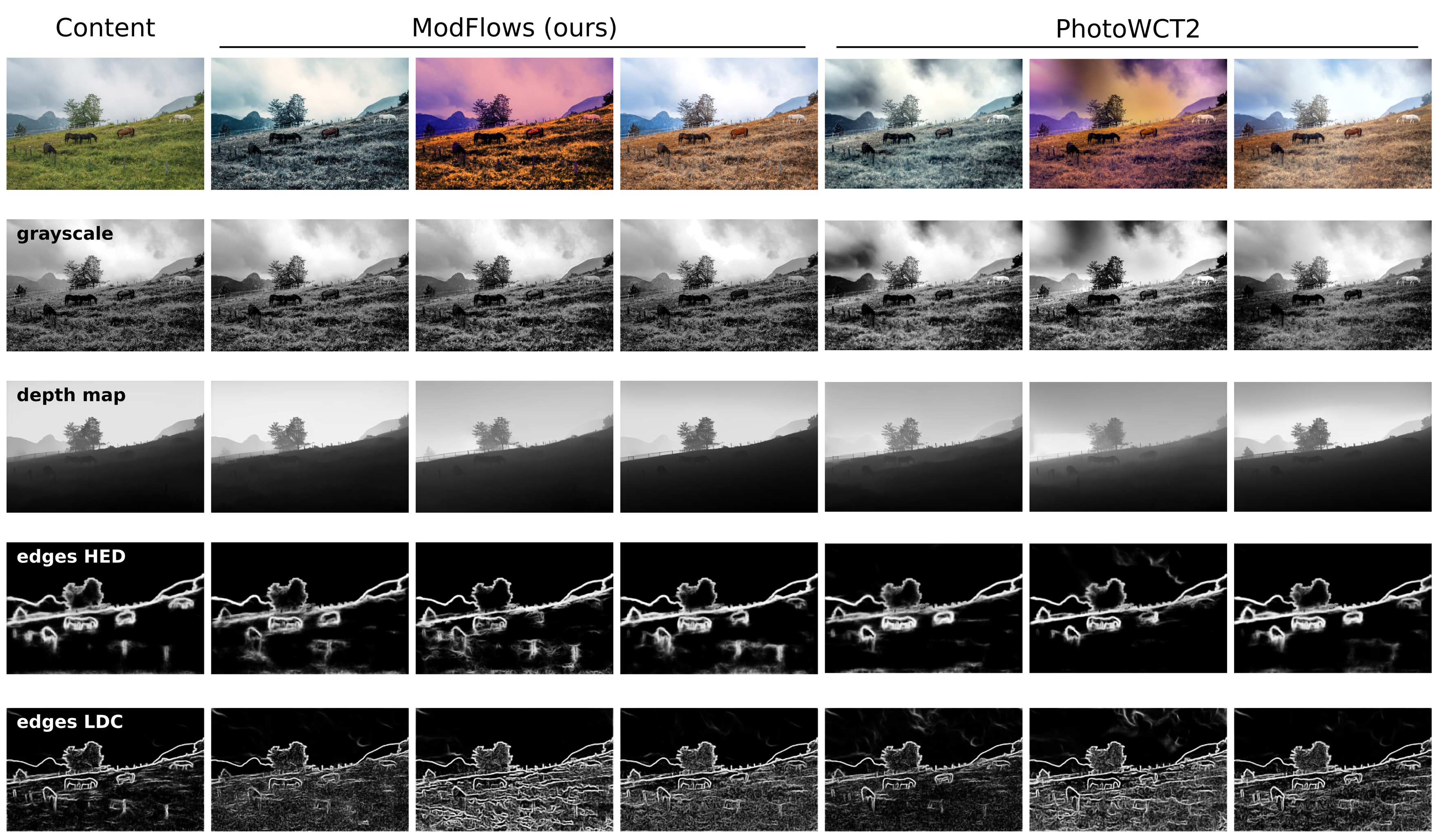


Figure: A content metric measures a structural similarity between a content image and a result. The choice of the best content metric is not obvious. Edges detection by HED model *(Xie et al. 2015)* grasp mostly the main objects of a scene, while LDC model *(Soria et al. 2022)* captures too detailed edges. Both of them are not sensitive to low-frequency artifacts. To show the absence of such artifacts in the Modflows we compute similarity scores between the depth maps *(Gui et al. 2024)* and the normalized grayscale images, which are processed to have a linear intensity histogram.

Contacts

GitHub

Maria A. Larchenko
mariia.larchenko@gmail.com

Alexander A. Lobashev
lobashevalexander@gmail.com