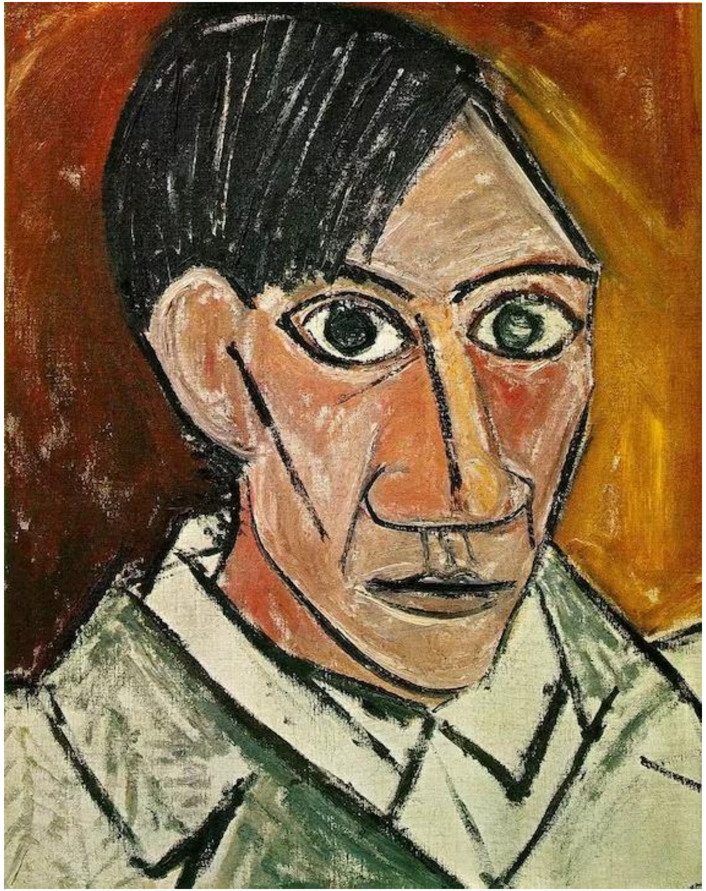
DIFF-NST: Style transfer with Diffusion models

Dan Ruta, Gemma Canet Tarrés, Andrew Gilbert, Eli Shechtman, Nicholas Kolkin, John Collomosse

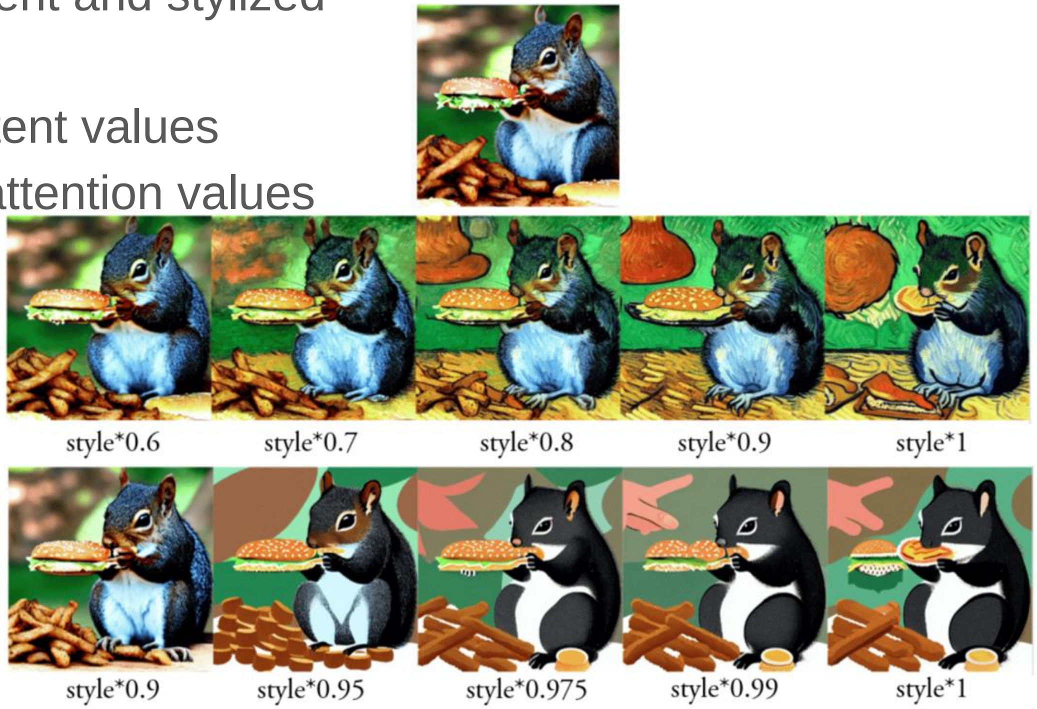
- Styles can focus on the form of its subject matter, rather than its rendering style
- This aspect of style hasn’t been explored as much in NST literature



DIFF-NST: Style transfer with Diffusion models

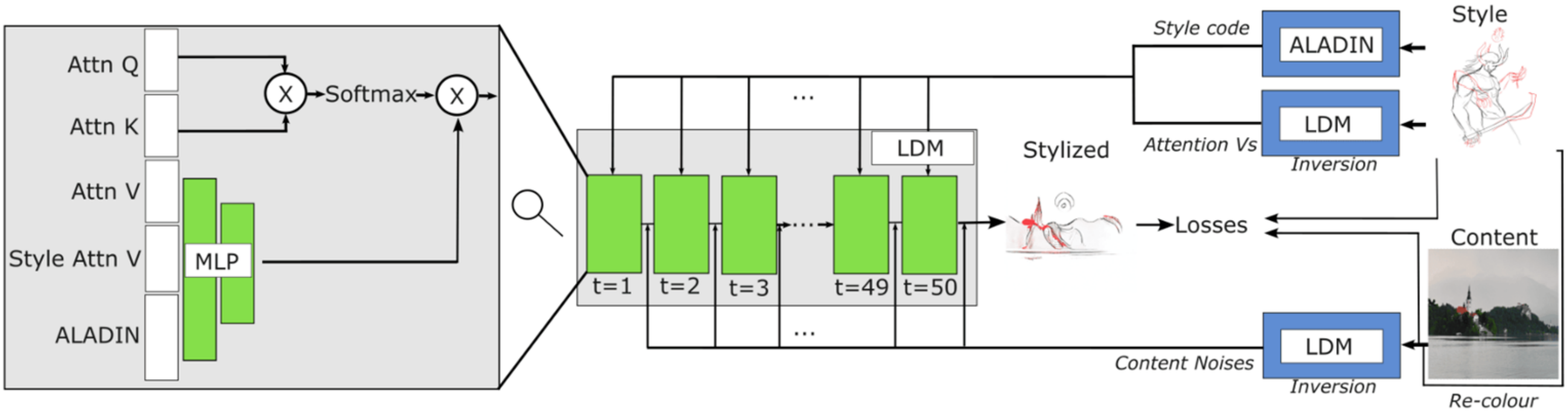
Preliminary experiments with prompt-to-prompt

- Generated many pairs of content and stylized images
- Analysed the differences in latent values
- Most changes occur in the V attention values



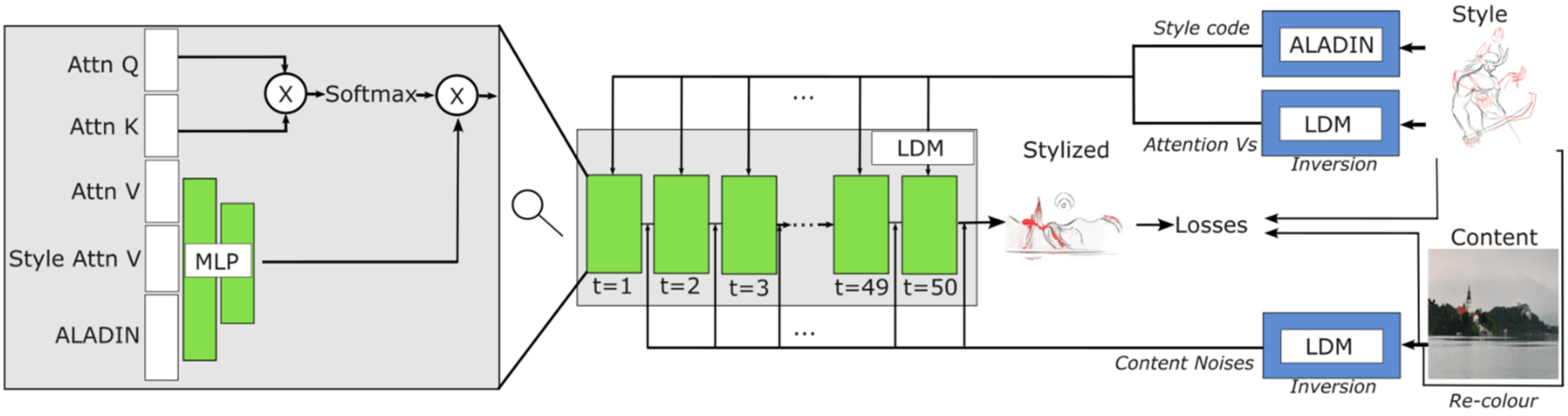
DIFF-NST: Style/Content inversion and injection

- Invert the content image through injecting the noises
- Inject the style information via the V attention values, and ALADIN style code



DIFF-NST: Unconditional generation

- Only unconditional branch is executed
- No text prompts are used at any point in the process



ALADIN-NST for avoiding semantic creep

- Strong disentanglement in ALADIN-NST avoids encoding strong semantic information such as faces into the stylization

Quantitative results

- Competitive results - despite the metrics not capturing high level deformation changes
- We score high on style ratings in user studies
- We score low on content ratings in user studies
    - Note, this is good, as we are not aiming to re-create the content details

Table 1: Quantitative metrics. Lower is better. ↓

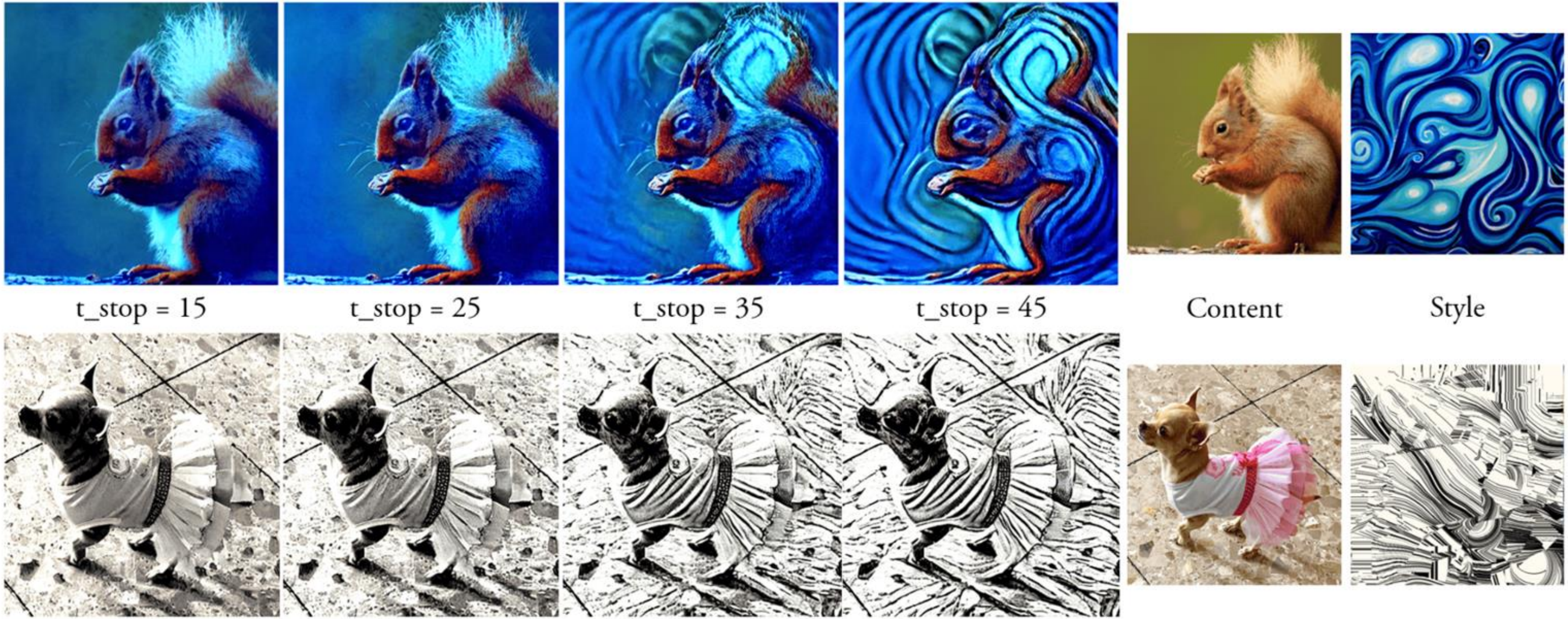
Model	LPIPS ↓	SIFID ↓	Chamfer ↓
NeAT [25]	0.624	0.880	24.970
CAST [41]	0.632	1.520	43.864
NNST [13]	0.633	2.007	53.328
PARASOL [31]	0.716	3.297	105.371
DIFF-NST (Ours)	0.656	2.026	45.777

Table 2: User studies for our model, for individual ratings (out of 5), and 5-way preferences (%). Higher is better. ↑

Model	Content Rating ↑	Style Rating ↑	Content Preference ↑	Style Preference ↑
NeAT [25]	3.271	2.952	32.222	26.000
CAST [41]	3.031	2.863	16.756	16.133
NNST [13]	2.937	2.712	21.200	17.778
PARASOL [31]	2.301	2.257	12.400	9.556
DIFF-NST (Ours)	2.751	2.973	17.422	30.533

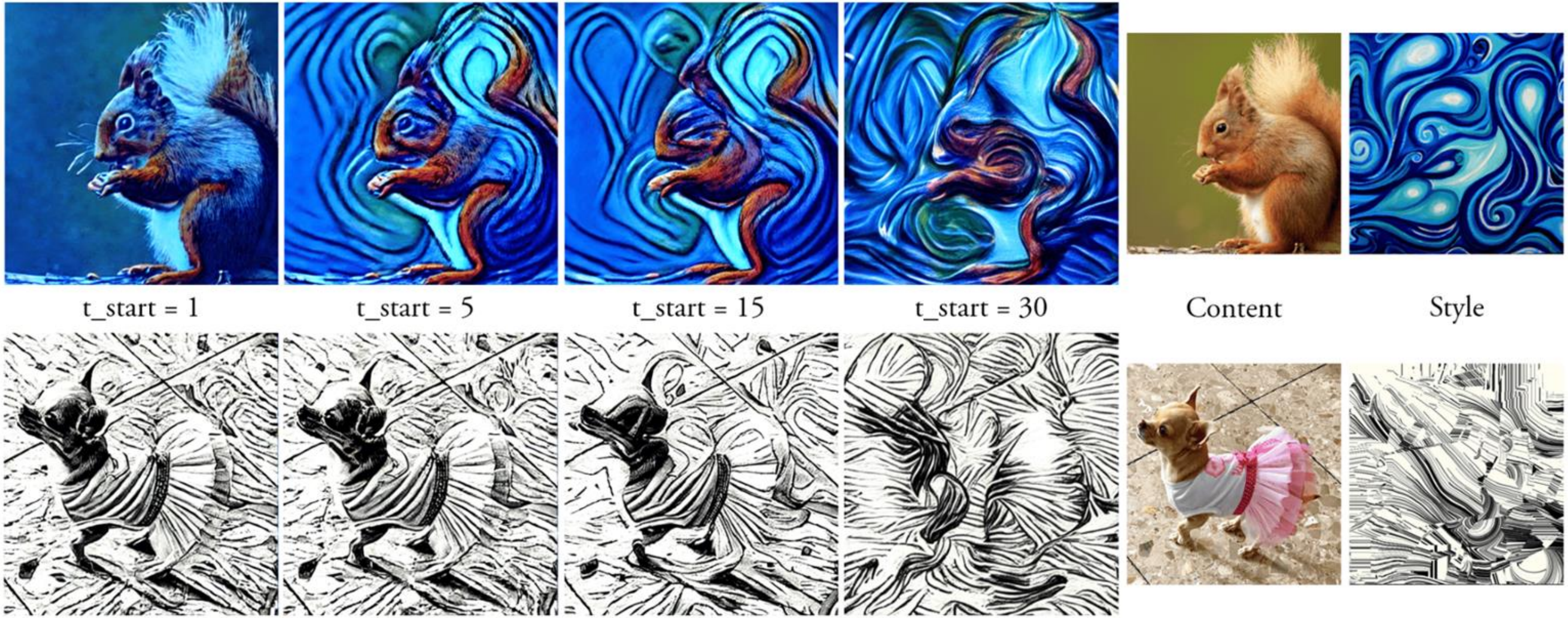
DIFF-NST stylization strength control

- Limiting the timestep until the style attention V and ALADIN code is



DIFF-NST deformation control

- Delaying the starting step for noise injection controls how much the style will

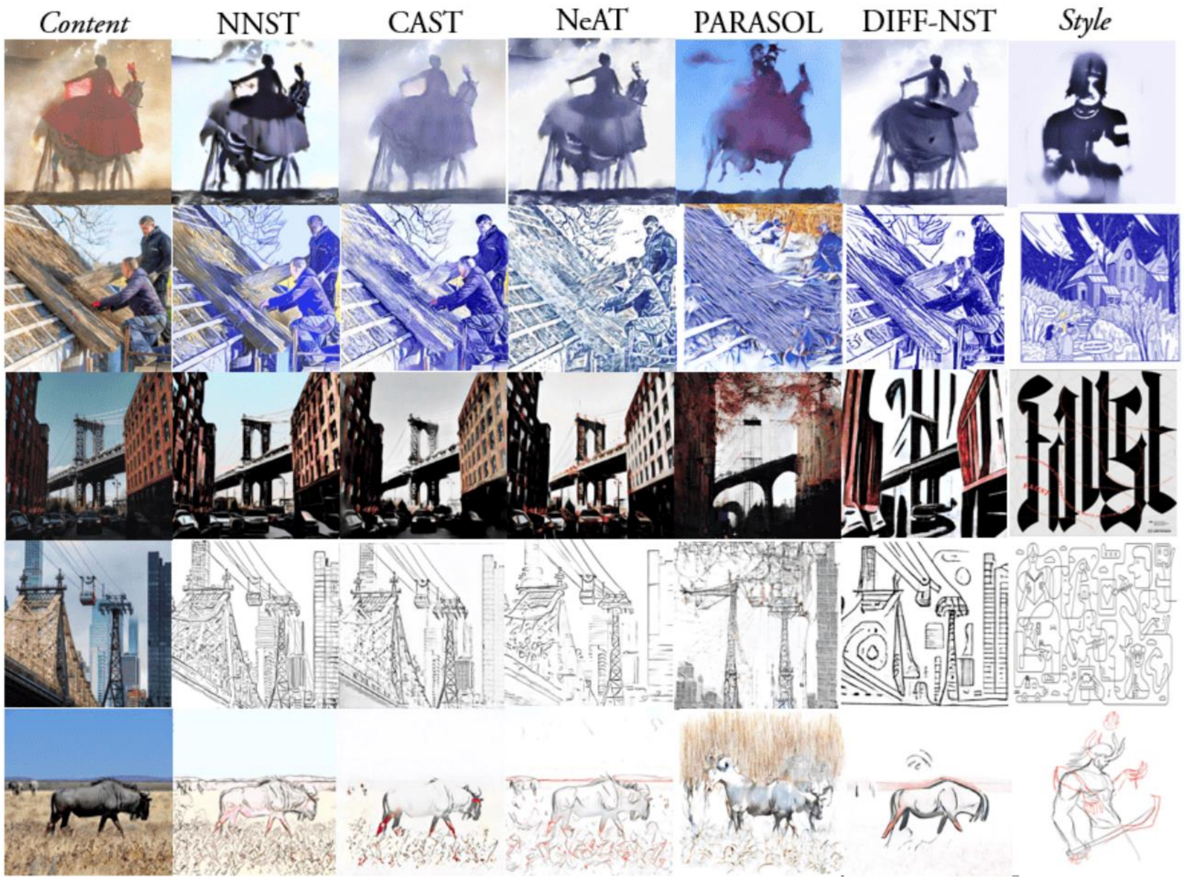


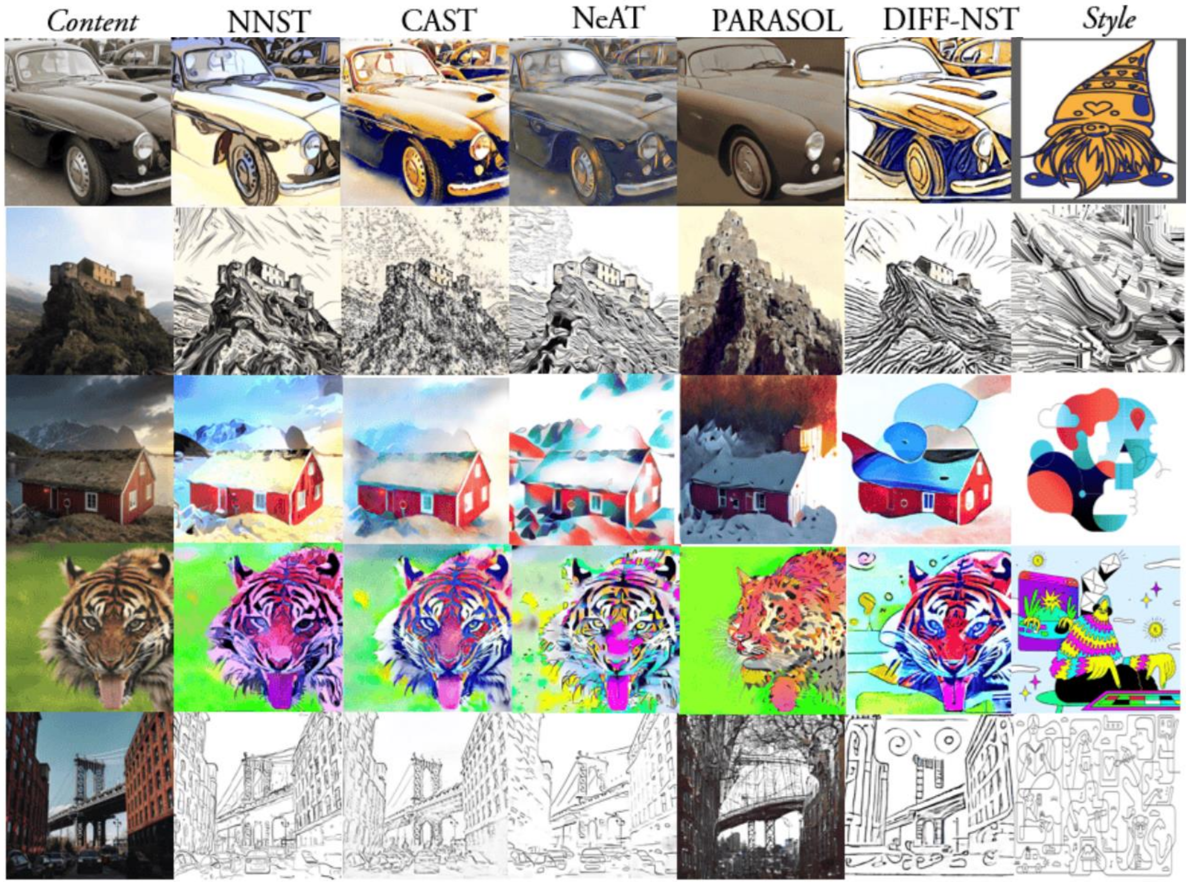
DIFF-NST: Style deformation

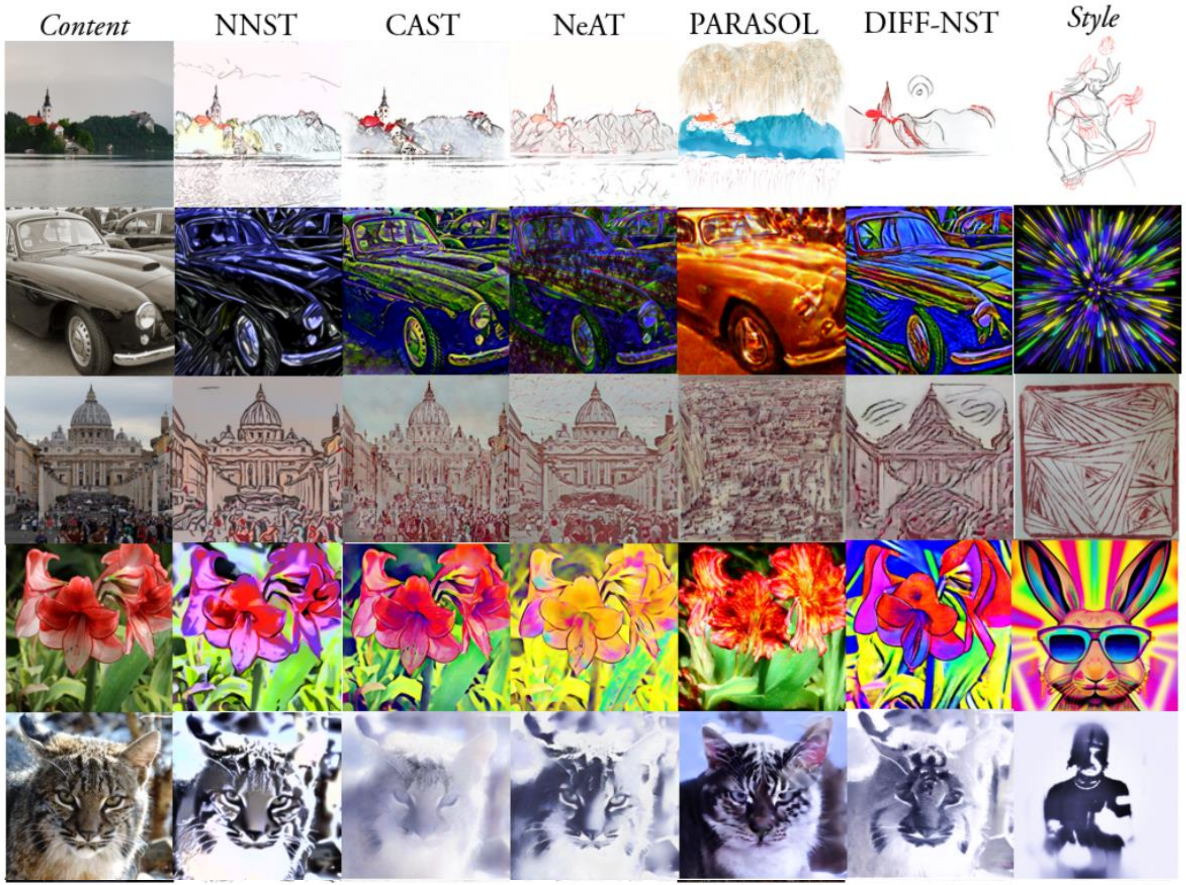
Conclusions

- Explored an application of NST with pre-trained diffusion models
- Enabled controllable deformation in NST for the first time
  - The abstraction of content can be controlled by the timestep of content information injection
  - Style strength can be controlled by the timestep of style information injection

Thank you


Content
NNST
CAST
NeAT
PARASOL
DIFF-NST
Style


Content
NNST
CAST
NeAT
PARASOL
DIFF-NST
Style


Content
NNST
CAST
NeAT
PARASOL
DIFF-NST
Style