고려대학교 빅데이터 학회

Style Transfer for Videos

알 고 리 즘 발 표

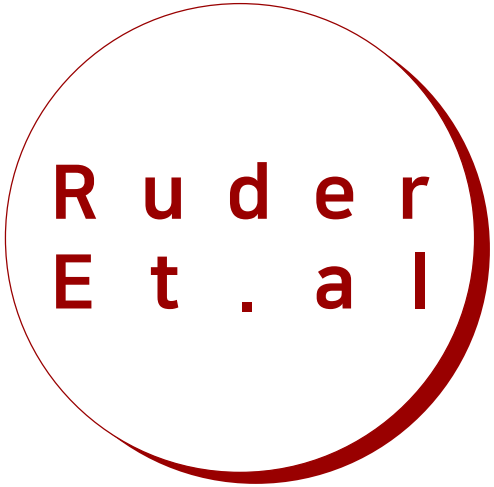
KU-BIG

읽었던 논문들

Ruder Et. al

ReCoNet

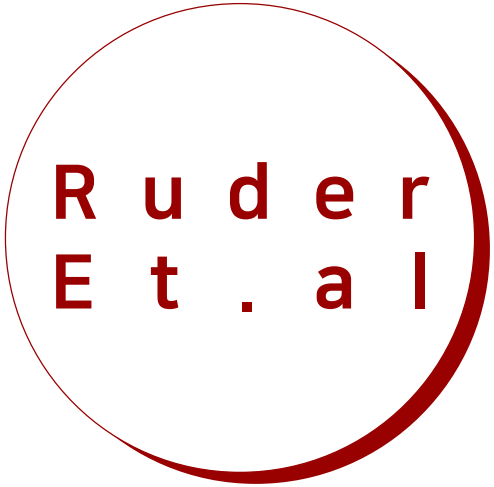
Artistic style transfer for videos



장점
- 성능이 매우 뛰어남

단점
- 컴퓨터 성능도 뛰어나야 됨
- 구현 난이도 ▲

Artistic style transfer for videos



ization the optimization process needed on average roughly eight to ten minutes per frame at a resolution of 1024 × 436 on an Nvidia Titan X GPU. When ini-

ReCoNet

ReCoNet Paper – Model & Notations

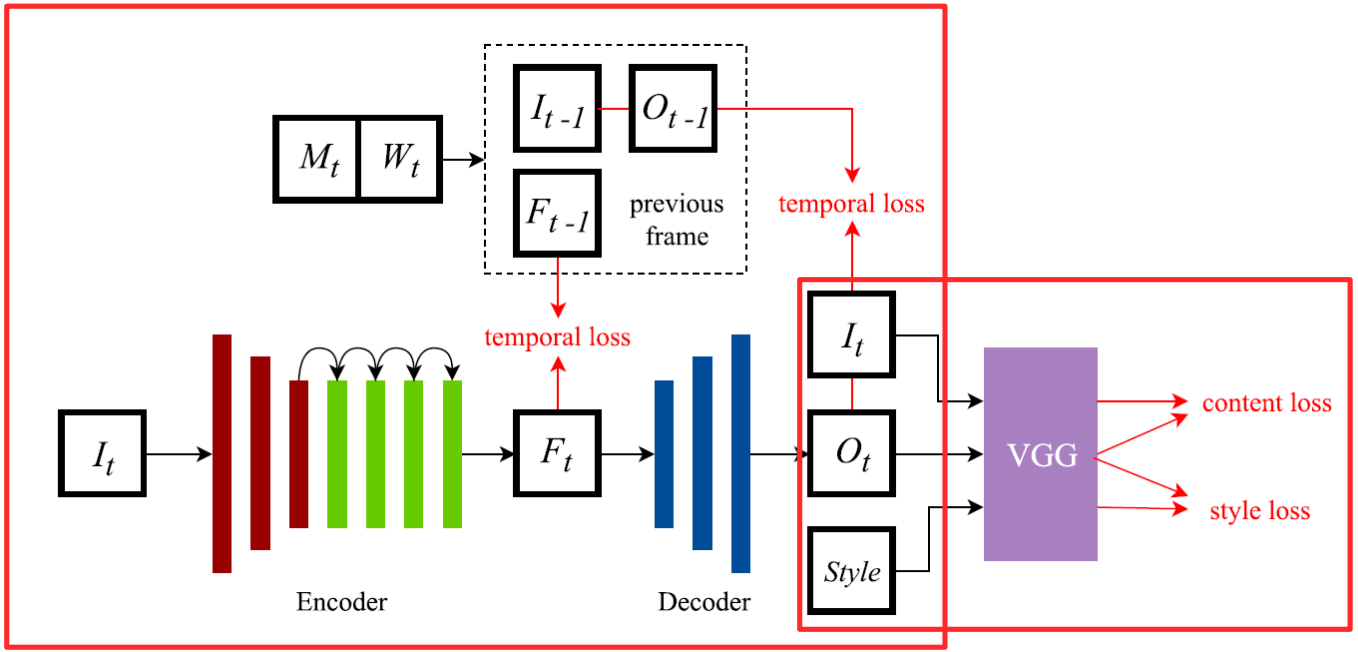


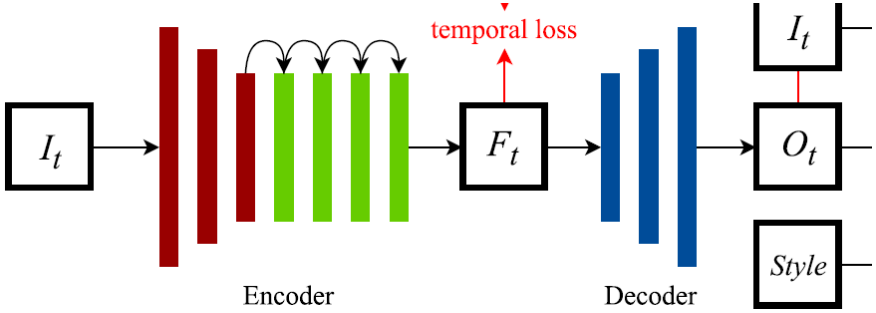
Image Style Transfer 와 유사한 부분

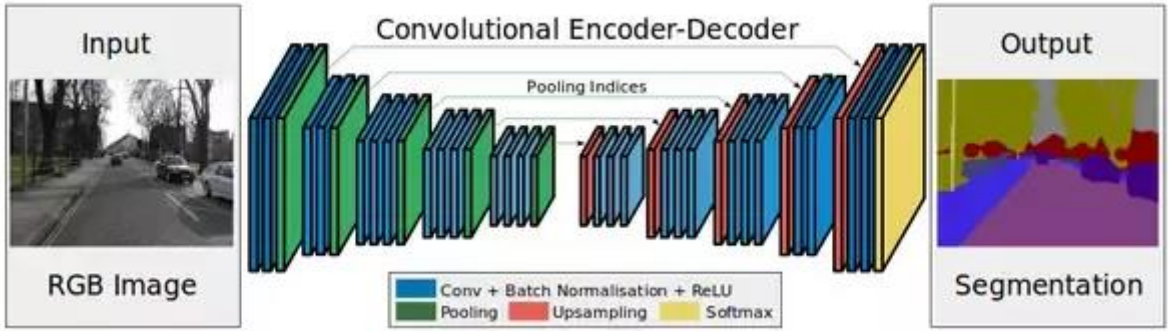
사전 지식

Encoder
&
Decoder

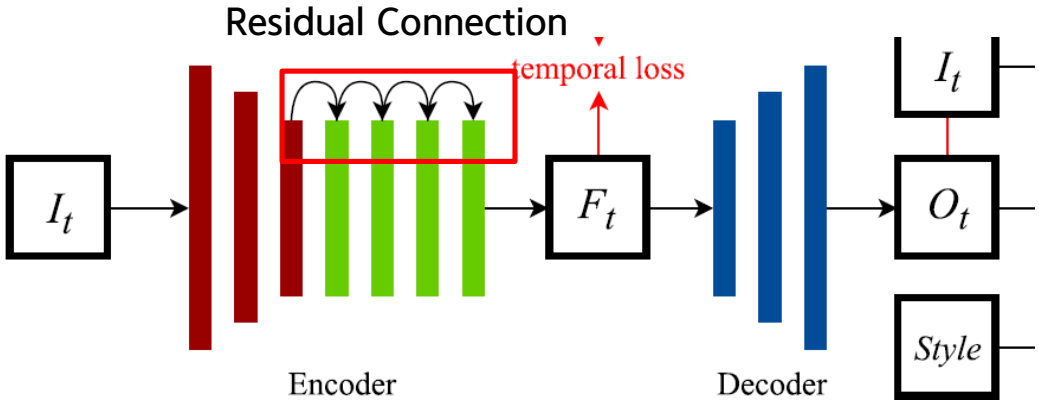
Optical
Flow

What is Encoder and Decoder?

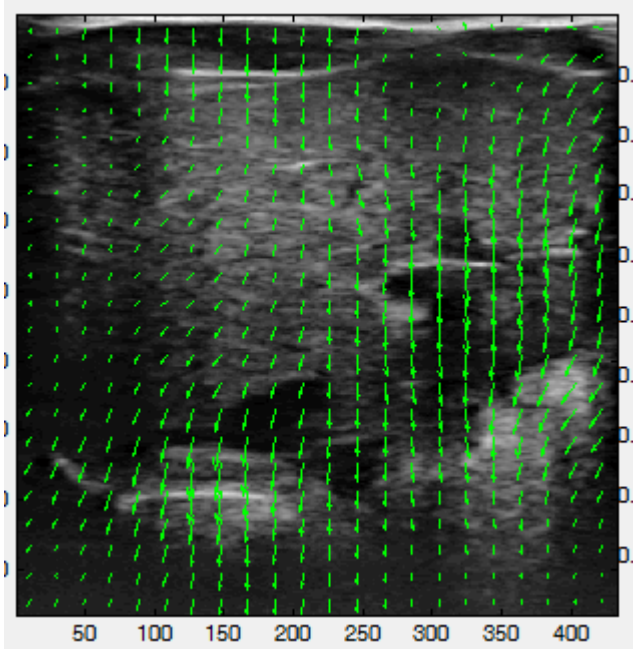




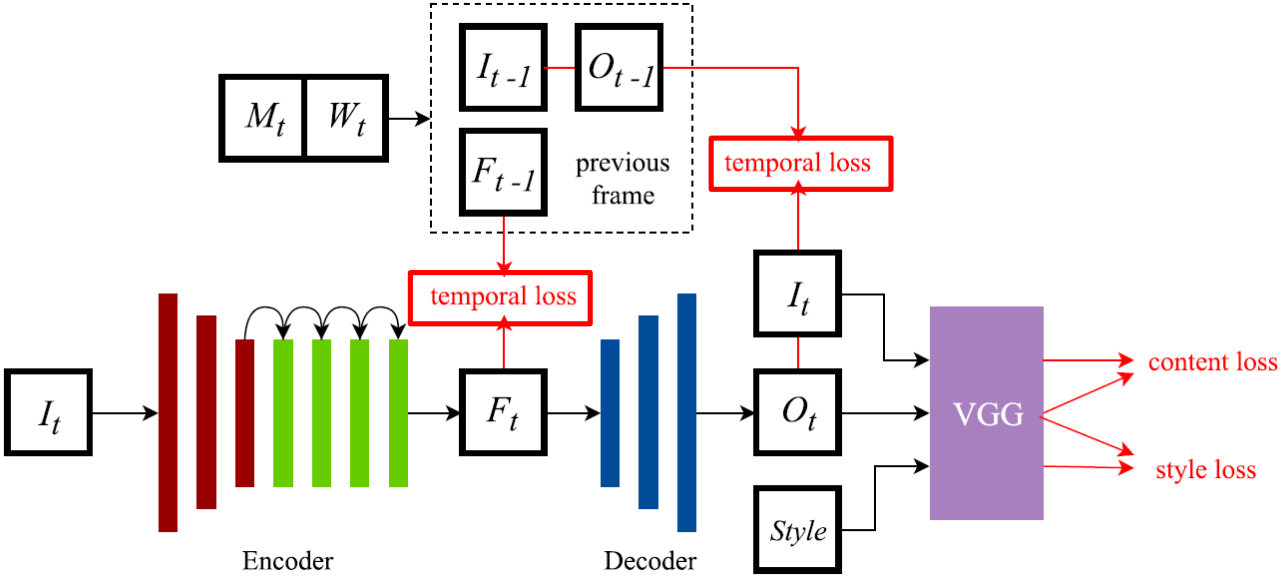
What is Encoder and Decoder?



Optical Flow & Occlusion Mask



ReCoNet Paper – Model & Notations



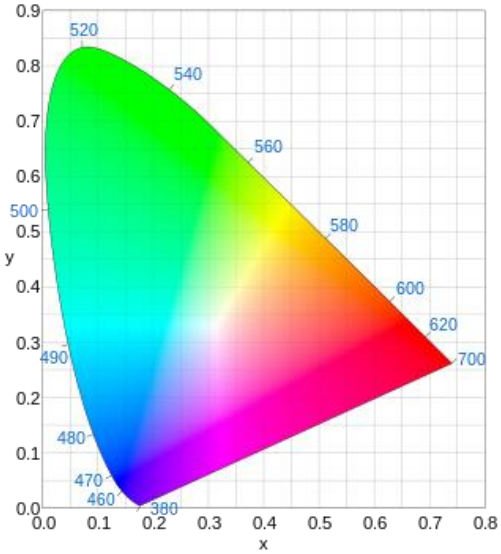
저자가 중요시 한 것

Y
Channel
Color

Temporal
L o s s

Y Channel Color

XYZ Color Space



Y(luminance)에 대해 집중함

What is Temporal Loss?

Video와 Picture을 다룰 때 가장 큰 차이?

What is Temporal Loss?

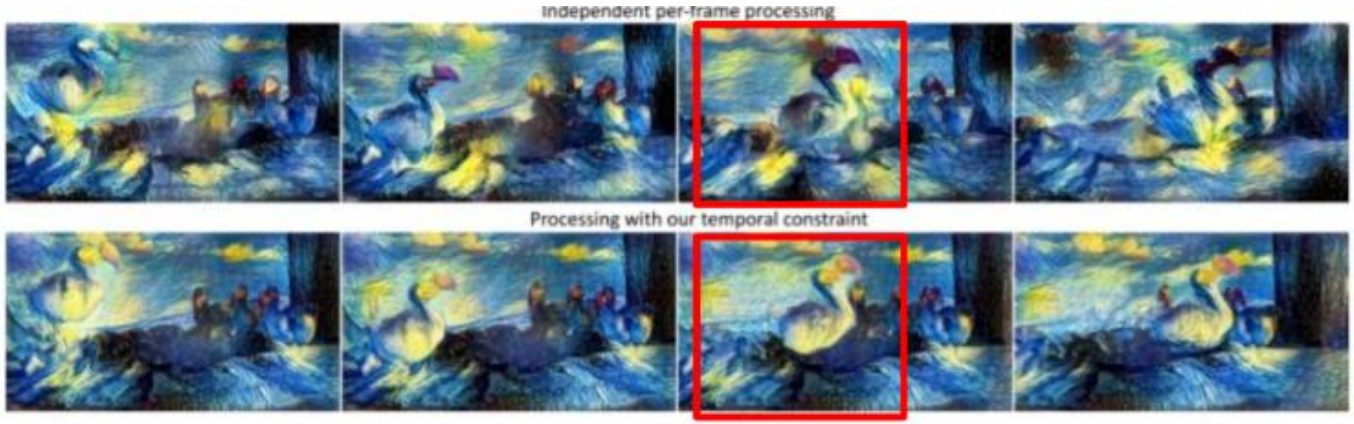
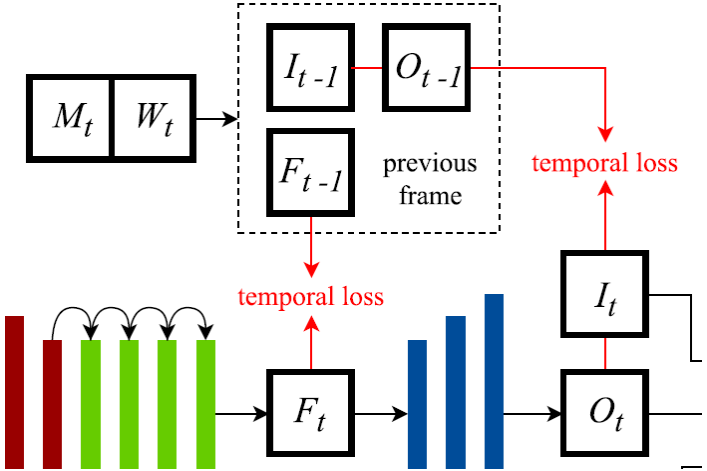


Fig. 1. Scene from *Ice Age* (2002) processed in the style of *The Starry Night*. Comparing independent per-frame processing to our time consistent approach, the latter is clearly preferable. Best observed in the supplemental video, see section 8.1.

Consistency

ReCoNet Paper – Model & Notations



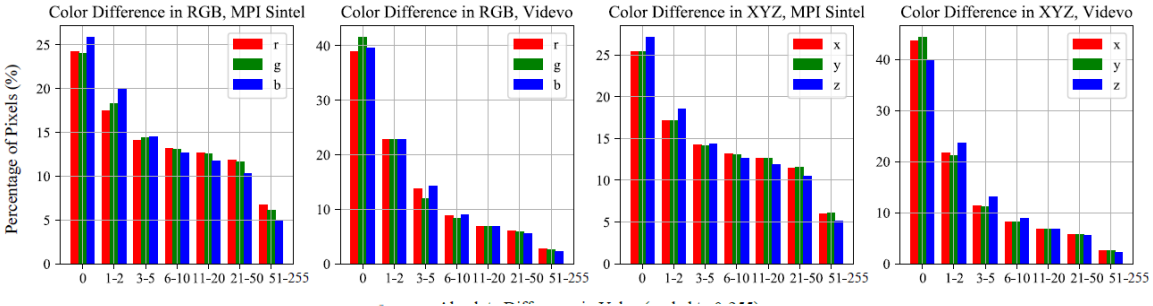
F_t: t번째 Frame의 Encoder Feature Map (output)
I_t: t번째 Frame의 Input
O_t: t번째 Frame의 Stylized Output
M_t: t번째 Frame의 Occlusion Mask
W_t: t-1 to t Frame 사이 Optical Flow

ReCoNet Paper – Model & Notations

Summary The final loss function for the two-frame synergic training is:

$$\mathcal{L}(t-1,t) = \sum_{i \in \{t-1,t\}} (\alpha \mathcal{L}_{content}(i) + \beta \mathcal{L}_{style}(i) + \gamma \mathcal{L}_{tv}(i)) + \lambda_f \mathcal{L}_{temp,f}(t-1,t) + \lambda_o \mathcal{L}_{temp,o}(t-1,t) \quad (3)$$

Output Level Temporal Loss



$$\mathcal{L}_{temp,o}(t-1,t) = \sum_{c} \frac{1}{D} M_t \| (O_t - W_t(O_{t-1}))_c - (I_t - W_t(I_{t-1}))_Y \|^2$$

I_t: t번째 Frame의 Input
O_t: t번째 Frame의 Stylized Output
M_t: t번째 Frame의 Occlusion Mask
W_t: t-1 to t Frame 사이 Optical Flow
D: 픽셀 개수

Feature Map Level Temporal Loss

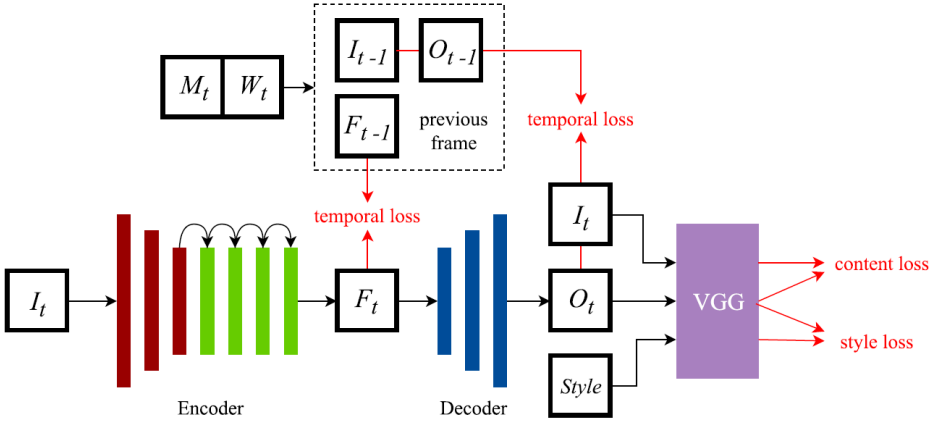
$$\mathcal{L}_{temp,f}(t-1,t) = \frac{1}{D} M_t \| F_t - W_t(F_{t-1}) \|^2$$

M_t: t번째 Frame의 Occlusion Mask
W_t: t-1 to t Frame 사이 Optical Flow
D: Feature Map에서 Element 개수

ReCoNet Paper – Model & Notations

Summary The final loss function for the two-frame synergic training is:

$$\mathcal{L}(t-1,t) = \sum_{i\in\{t-1,t\}} (\alpha\mathcal{L}_{content}(i) + \beta\mathcal{L}_{style}(i) + \gamma\mathcal{L}_{tv}(i)) + \lambda_f\mathcal{L}_{temp,f}(t-1,t) + \lambda_o\mathcal{L}_{temp,o}(t-1,t) \quad (3)$$



ReCoNet의 한계

장점
- 구현한 코드가 존재
- 속도가 제일 빠름

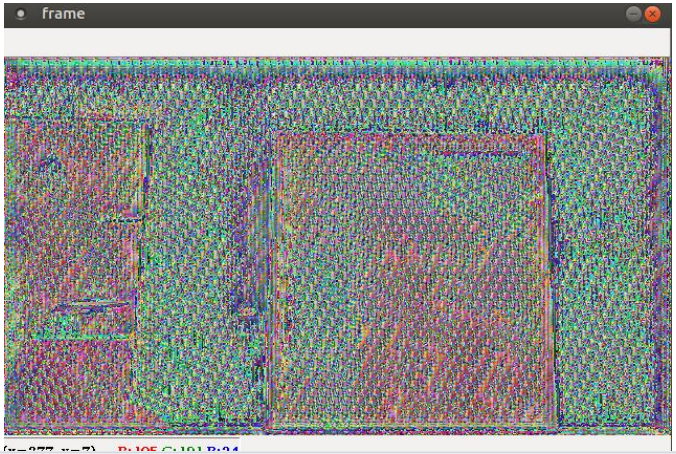
단점
- 그 구현이 심각함
- 성능이 SOTA가 아닌 듯 함

ReCoNet의 단점

Table 2. Temporal error e_{stab} and average FPS in the inference stage with style *Candy* on different models. Five scenes from MPI Sintel Dataset are selected for validation

Model	Alley-2	Ambush-5	Bandage-2	Market-6	Temple-2	FPS
Chen *et al* [4]	0.0934	0.1352	0.0715	0.1030	0.1094	22.5
ReCoNet	0.0846	0.0819	0.0662	0.0862	0.0831	235.3
Huang *et al* [17]	0.0439	0.0675	0.0304	0.0553	0.0513	216.8
Ruder *et al* [27]	0.0252	0.0512	0.0195	0.0407	0.0361	0.8

ReCoNet의 단점 2



xinix909 commented 3 days ago

@justanhduc @AntoineGerardeaux @Yiman-GO I fixed it by delete a "mul(255)" or something, there is more other bugs in the code. And now, still no temporal consictency. Flow wrap method is wrong I think.

Huang의 구현체는?

Commits on Feb 7, 2018

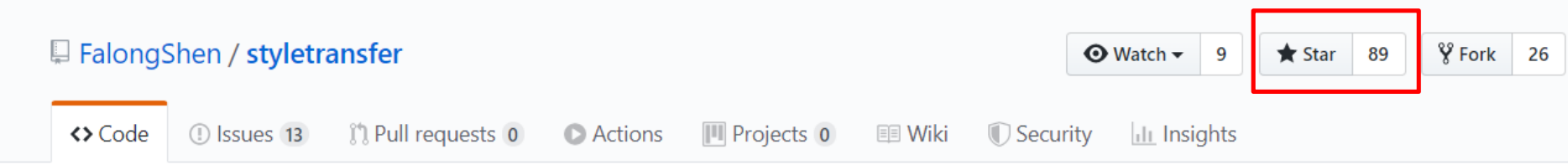
transfer network 완성, loss나 optical flow가 맞는지는 모름

 curaai00 committed on 7 Feb 2018

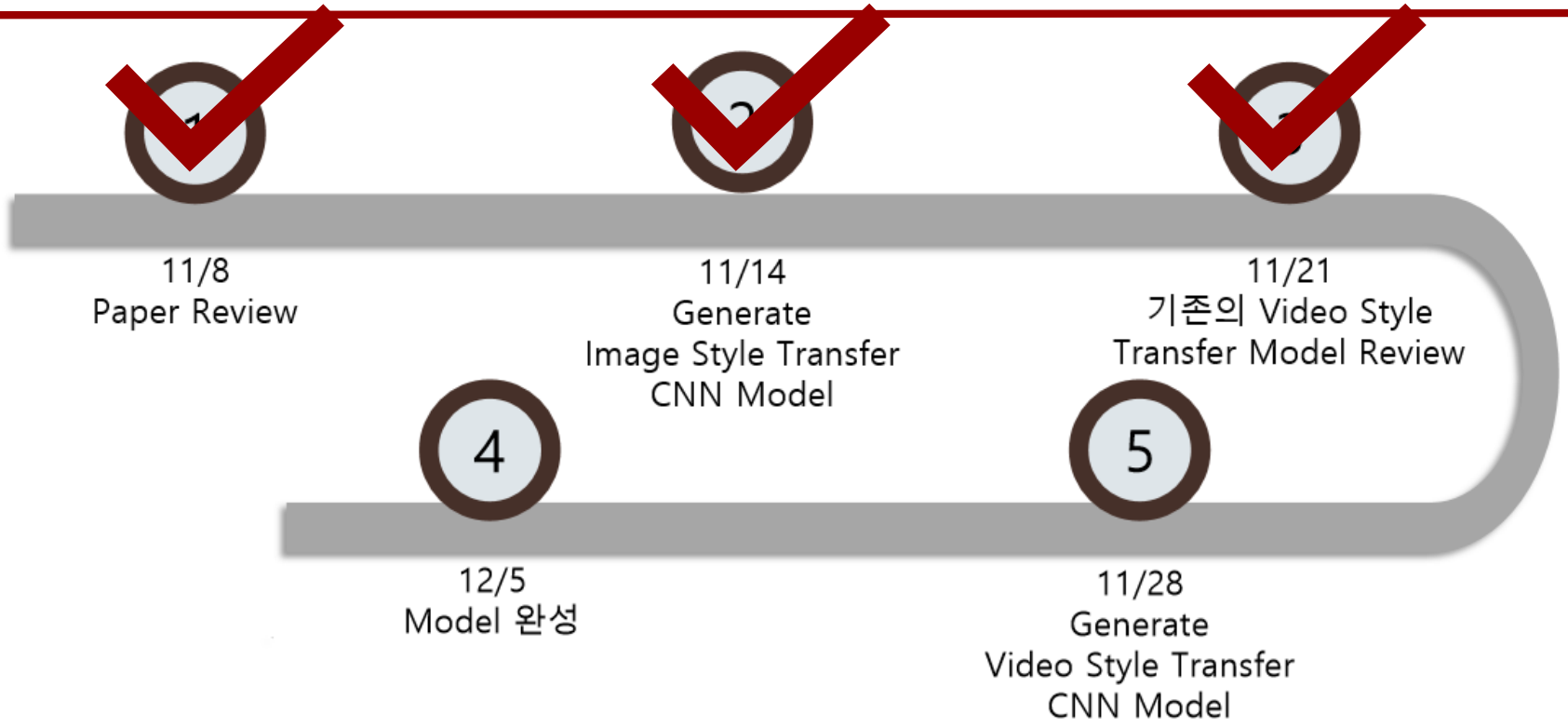
Dataset, Video crawler(videvo.net)

 curaai00 committed on 7 Feb 2018

Huang의 구현체는?


FalongShen / styletransfer
Watch 9
Star 89
Fork 26
Code
Issues 13
Pull requests 0
Actions
Projects 0
Wiki
Security
Insights
Real-time neural style transfer via meta networks https://github.com/FalongShen/styletr...
Installation
This library is based on Caffe. CuDNN 7 and NCCL 1 are required. Please follow the installation instruction of Caffe.

Further plans



Thank you