NYU

AI Calligraphic Poet

By: Haoqi (Kevin) Zhang, Millie Chen, Yuanheng Li

GitHub:
https://github.com/AlanMuErDan/DS-UA-301-AI-Calligraphic-Poet-Project/tree/main

Executive Summary – Problem Statement/Goal

Model 1 | Model 2 | Model 3

一个湖上，有许多红色的荷花和很多绿色的荷叶

On a lake, there are many red lotus flowers and many green lotus leaves

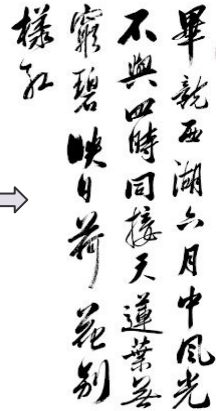
毕竟西湖六月中，风光不与四时同。
接天莲叶无穷碧，映日荷花别样红。

Now it comes, mid June on West Lake,

Four seasons, the vista ever unique.

Lotus leaves to the horizon, boundless green,

Sun glow on lotus buds, peerless red.



1. Images

2. Captions

3. Poems

4. Calligraphy

Our Approach: Image Captioning => Poem Gen => Calligraphy Gen

NYU

Executive Summary – Challenges, Value

Challenges:

1. Generation, not classification or regression
2. Multimodal, cross-sectional (CV, NLP)

Value/Benefit:

Combining photography, poetry, calligraphy into a unified application

Related Works/ Motivation

1. For image captioning: GIT and other models focus on English captions ➡ our target language is Chinese
2. For Poem generation: SongNet ➡ Tokenization & Output
3. For calligraphy generation: GAN ➡ paired data

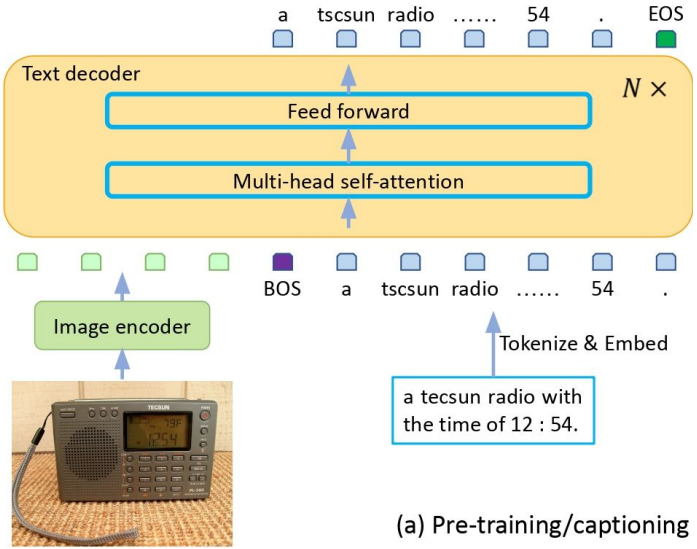
Motivation:

We believe it's innovative to bring together different forms of art in a creative and dynamic way

NYU

Method/Approach

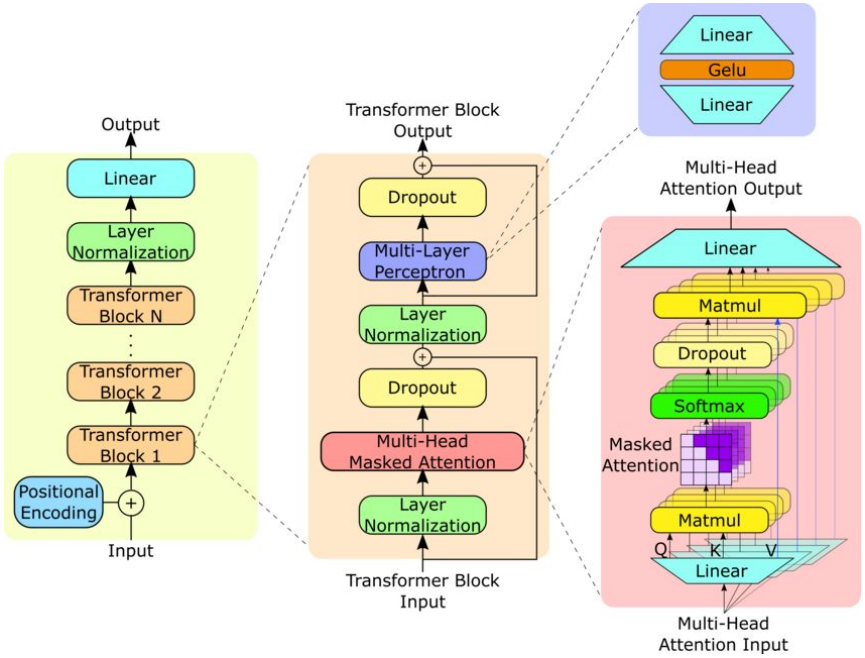
GIT: Generative Image to Text Transformer

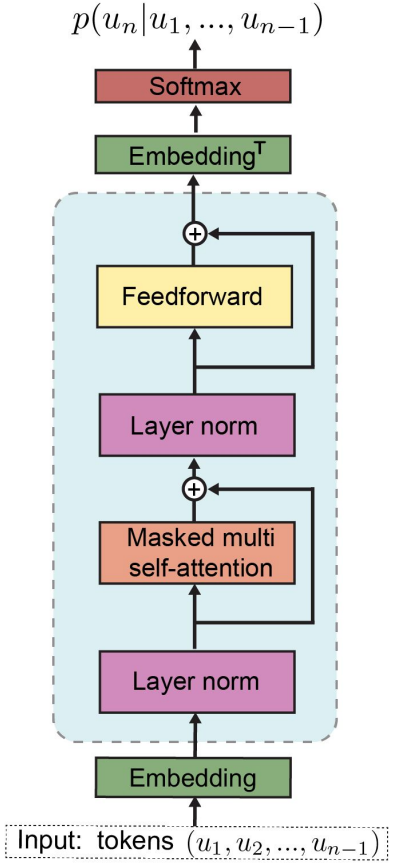


Source: https://arxiv.org/abs/2205.14100

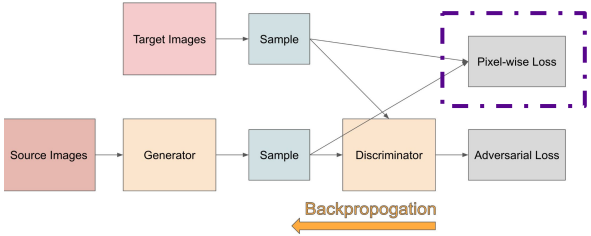
Method/Approach

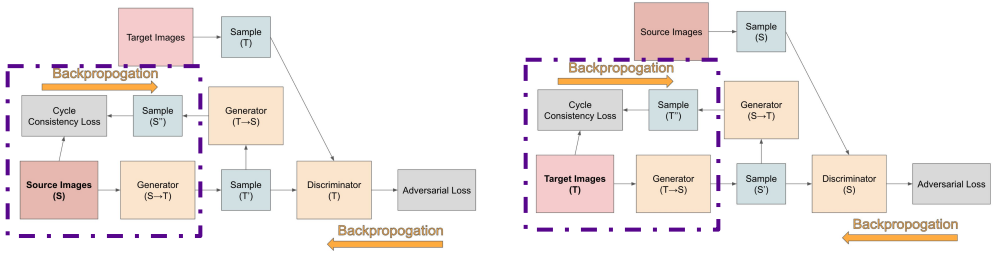
Fine-tuning: GPT 2





Method/Approach





$$\mathcal{L}_{\text{adv}} = E_x \left[\log D(x)\right] + E_z \left[\log(1 - D(G(x)))\right]$$

$$\mathcal{L}_{\text{pixel}} = \|G(x) - y\|_1$$

$$\mathcal{L} = \mathcal{L}_{\text{adv}} + \lambda \mathcal{L}_{\text{pixel}}$$

$$\mathcal{L}_{\text{adv}}(G_{X\to Y}, D_Y) = \mathbb{E}_{y\sim Y}\left[\log(D_Y(y))\right] + \mathbb{E}_{x\sim X}\left[\log(1 - D_Y(G_{X\to Y}(x)))\right]$$

$$\mathcal{L}_{\text{adv}}(G_{Y\to X}, D_X) = \mathbb{E}_{x\sim X}\left[\log(D_X(x))\right] + \mathbb{E}_{y\sim Y}\left[\log(1 - D_X(G_{Y\to X}(y)))\right]$$

$$\mathcal{L}_{\text{cycle}}(G_{X\to Y}, G_{Y\to X}) = \mathbb{E}_{x\sim X}\left[\|G_{Y\to X}(G_{X\to Y}(x)) - x\|_1\right] + \mathbb{E}_{y\sim Y}\left[\|G_{X\to Y}(G_{Y\to X}(y)) - y\|_1\right]$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{adv}}(G_{X\to Y}, D_Y) + \mathcal{L}_{\text{adv}}(GY \to X, D_X) + \lambda \mathcal{L}_{\text{cycle}}$$

NYU

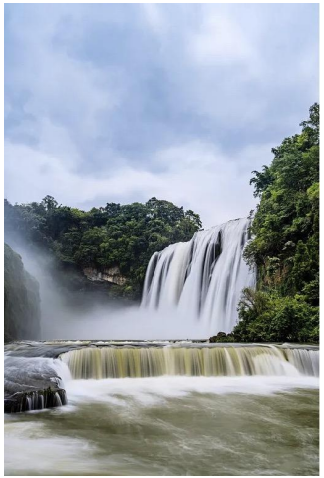
GAN

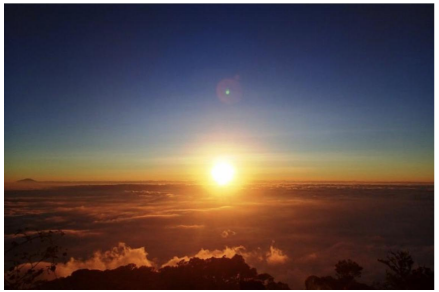
CycleGAN

Implementation/Experimentation Details

Methods

- New Tokenizer
- Fine Tuning layers related to caption generation

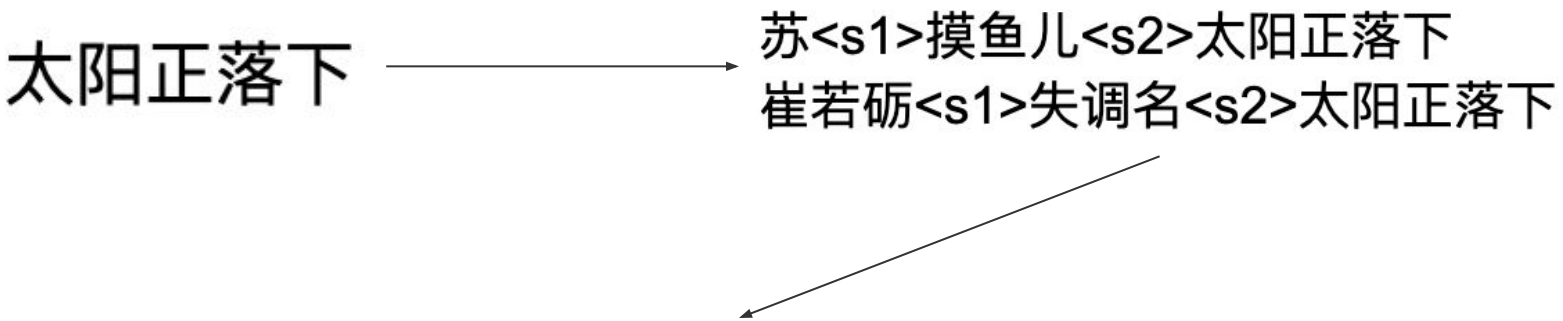




Current Dataset Examples

NYU

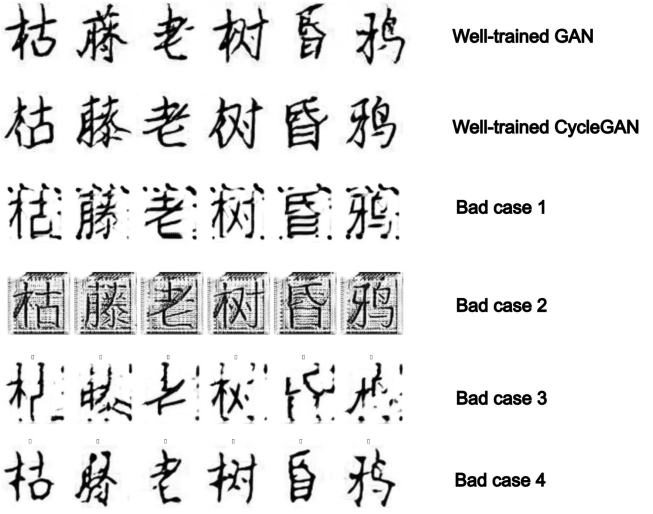
Implementation/Experimentation Details



苏<s1>摸鱼儿<s2>太阳正落下 <bos>倚棹三生惭</s>
苏<s1>摸鱼儿<s2>太阳正落下 <bos>是非都说。</s>
苏<s1>摸鱼儿<s2>太阳正落下 <bos>甚时是歇。</s>
苏<s1>摸鱼儿<s2>太阳正落下 <bos>枕上醉乡。</s>
苏<s1>摸鱼儿<s2>太阳正落下 <bos>也解忆著。</s>
苏<s1>摸鱼儿<s2>太阳正落下 <bos>这回归去住</s>
崔若砺<s1>失调名<s2>太阳正落下 <bos>失真丹方。</s>
崔若砺<s1>失调名<s2>太阳正落下 <bos>失马蹄筌。</s>
崔若砺<s1>失调名<s2>太阳正落下 <bos>失惊相将。</s>
崔若砺<s1>失调名<s2>太阳正落下 <bos>失却前踪。</s>
崔若砺<s1>失调名<s2>太阳正落下 <bos>失脚端阳。</s>
崔若砺<s1>失调名<s2>太阳正落下 <bos>失诚何德。</s>

NYU

Implementation/Experimentation Details



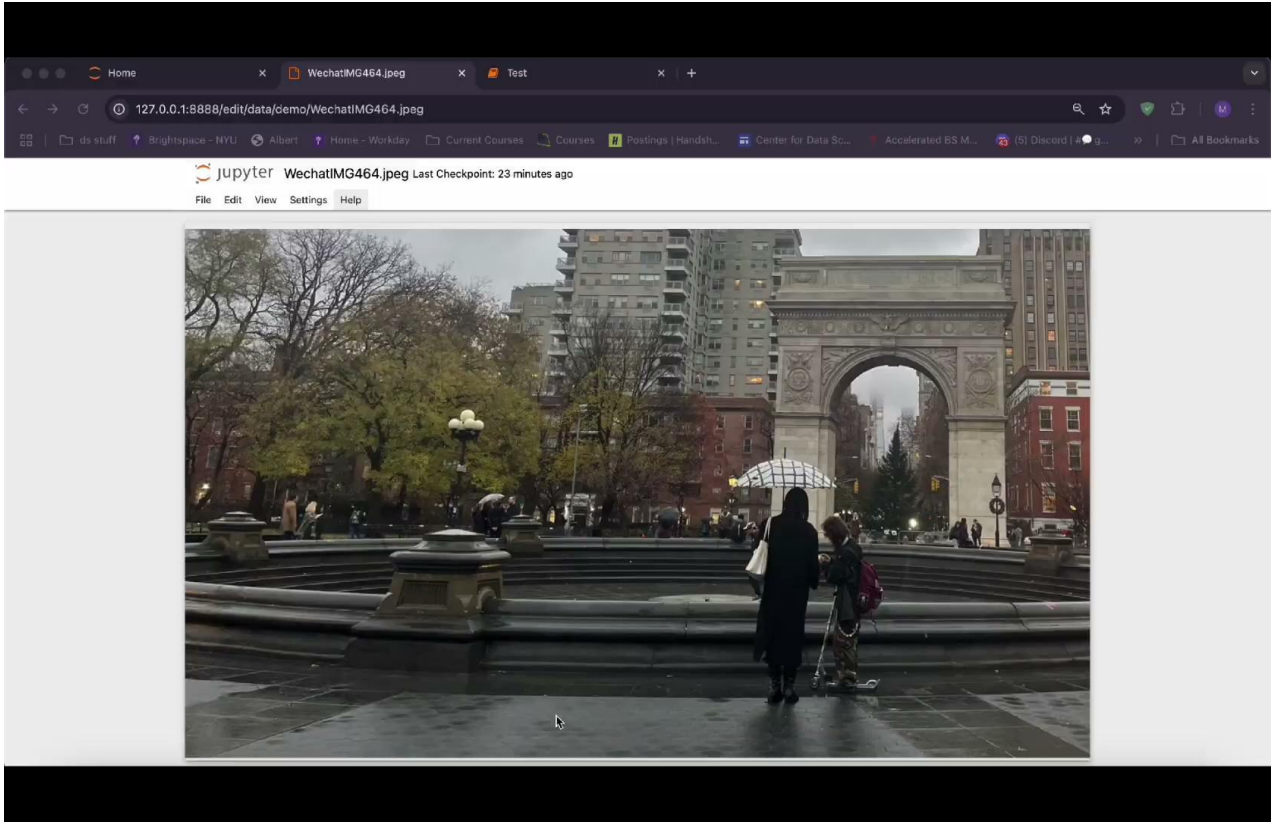
- Adding noise in **Discriminator**
- Dense Block (5) in **Generator**
- Leaky ReLU (0.2)
- Small lr = 0.0002
- Lambda = 10

$$\mathcal{L} = \mathcal{L}_{\text{adv}} + \lambda \mathcal{L}_{\text{pixel}}$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{adv}}(G_{X \to Y}, D_Y) + \mathcal{L}_{\text{adv}}(GY \to X, D_X) + \lambda \mathcal{L}_{\text{cycle}}$$

NYU

Recorded Demo



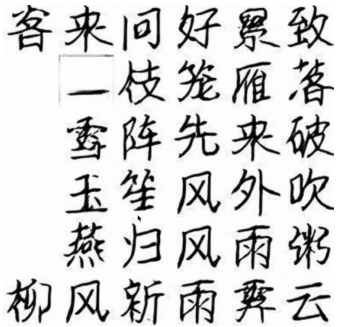
NYU

Results/Experimental Evaluation

天空是阴云的
The sky is cloudy

客来问好景致，一枝笼雁落。
雪阵先来破，玉笙风外吹。
燕归风雨粥，柳风新雨霁云。

A guest arrives, asking of the splendid scene,
A single branch holds a resting wild goose serene.
Snowstorms break first, the jade flute sings,
Its tune carried softly on windswept wings.
Swallows return through wind and rain's embrace,
Willows stir as new clouds clear, and rains erase.

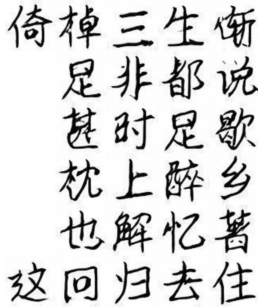


Results/Experimental Evaluation

太阳正落下
The sun is setting

倚棹三生惭，是非都说。
甚时是歇，枕上醉乡。
也解忆著，这回归去住。

Leaning on the oar, a lifetime's shame weighs,
Right and wrong fade with the passing days.
When will it end, this restless quest?
In drunken dreams, I long for rest—
To return and dwell where my heart lays.



远处的山
Mountain in the afar

朝复暮复，春山叠。
花飞时，闲窗翦。
裙裥皱，宫纱布。

Morning to dusk, the spring mountains rise,
Petals drift down as the flowers fly.
By the quiet window, time flows slow,
Folds in the skirt, soft gauze in tow,
The mountains linger beneath endless skies.

NYU

Results/Experimental Evaluation

Model	Normal	CSL	FWQ
VGG16	0.9407	0.6487	0.7731
ResNet50	0.9963	0.9493	0.9914
DenseNet-121	0.9997	0.9755	0.9931
Inception-v3	1.0000	0.9216	0.9851

Table 1: Classification Accuracy for Different CNN Architecture

Model	Style Score	OCR accuracy
GAN	0.8790	68%
CycleGAN	0.8732	69.2%

Table 2: Performance for GAN and CycleGAN on *bdsr* style

	BLEU-1	BLEU-2	METEOR
GIT	84	69.8	31.9
Our Model	36	24	40

Table 3: Performance of our finetuned model compared to original GIT model

Input	**Generated Text**	**Score**	**Comment**
苏<s1>摸鱼儿<s2>太阳正落下。</s>	<bos>又送一番沪。</s>	0.2927	Best match
崔若砺<s1>失调名<s2><太阳正落下。</s>	<bos>失信南枝晓。</s>	0.1527	Moderate match
崔若砺<s1>失调名<s2>太阳正落下。</s>	<bos>失诚恳悔别。</s>	0.1476	Low match
苏<s1>摸鱼儿<s2>太阳正落下。</s>	<bos>问予能几岁。</s>	0.1470	Moderate match
崔若砺*s1*失调名<s2>太阳正落下。</s>	<bos>失前村路古。</s>	0.1477	Low match
崔若砺<s1>失调名<s2>太阳正落下。</s>	<bos>失脚浅胜雪。</s>	0.1540	Moderate match
苏<s1>摸鱼儿<s2>太阳正落下。</s>	<bos>倚醉君休辞。</s>	0.1468	Low match

Table 1: Experimental results showing input, generated text, score

Conclusion

Discussion & Conclusions:

1. It would be a cool application
2. Hard to make a sophisticated product

Future works:

1. Pipeline
2. Expanding the scope
3. Build Frontend (UI/application/website)

Thank You