

AI-5 Course Project

Perception.AI

Presented by Team Zenith

Anshika, Harsh, Meghana, Vishnu →

Project Outline

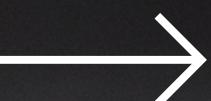
Problem Definition

Proposed Solution

Project Scope

Dataset Details

Model considerations



Problem Definition

Automatic image generation from captions is a challenging current areas of research, at the forefront of efforts in both NLP and Vision work. High performance on this task would have a number of practical applications. One example would be image search: a user may type a complicated caption phrase, a model for this task can generate an image matching this caption, and a reverse image search based on the content of the generated image can be used to find relevant existing results.



Proposed Solution



We propose to build an Adversarial Network (GANs) model for generating images from captions. Specifically, we focus our efforts on improving model performance and overall quality of generated by experimenting with various GAN architectures and language models. We deploy the best work in the form of a user application to take text or speech input for desired image generation.

Project Scope

/05

Proof Of Concept (POC)

- Creating a baseline model, maybe a StackGAN for text-to-image synthesis.
- Using subsets of CUB bird dataset and COCO dataset for baseline models and comparing GANs.
- Verifying by generating new images.



Prototype

- Create a mockup of screens to see how the app could look like.
- Deploy one model to Fast API to service model predictions as an API.



Minimum Viable Product (MVP)

- Create App to generate images from text or speech.
- API Server for uploading text / speech and generating images using best model.



Dataset Details

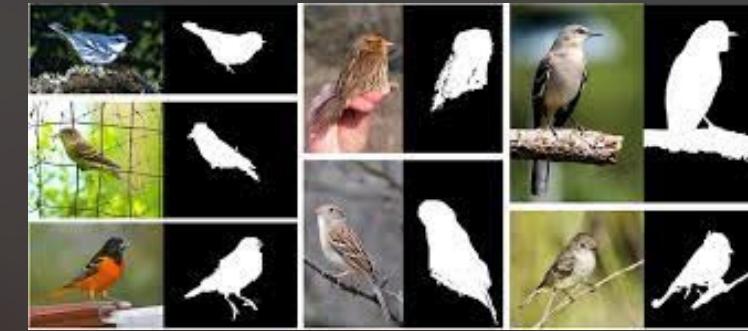
/06



COCO dataset

- We will use the subset of the latest release for the baseline model.
- The latest release has 118K images for training and 5K for validation.

CUB-200- 2011



- 11,788 images of 200 subcategories belonging to birds
- widely-used dataset for fine-grained visual categorization task.

Oxford 102 Flower

Misclassification	Test Image	English Marigold	Tree Poppy	Hibiscus
		Barberton Daisy		
		Japanese Anemone		
		Azalea		

- image classification dataset consisting of 102 flower categories.
- Each class consists of between 40 and 258 images.



Models

Stack GANs ++

We propose Stacked Generative Adversarial Networks (StackGAN) aiming at generating high-resolution photo-realistic images, as one of the baseline and comparative models.



IMAGAN

We propose this model for taking in text inputs and generating desired images.

