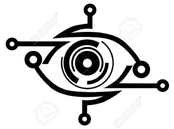
Perception.AI



Deploying Text to Image Generative Model

Anshika, Harsh Vardhan, Meghana and Vishnu

Problem Statement

Text to Image

Input = Textual description (e.g. Image caption)

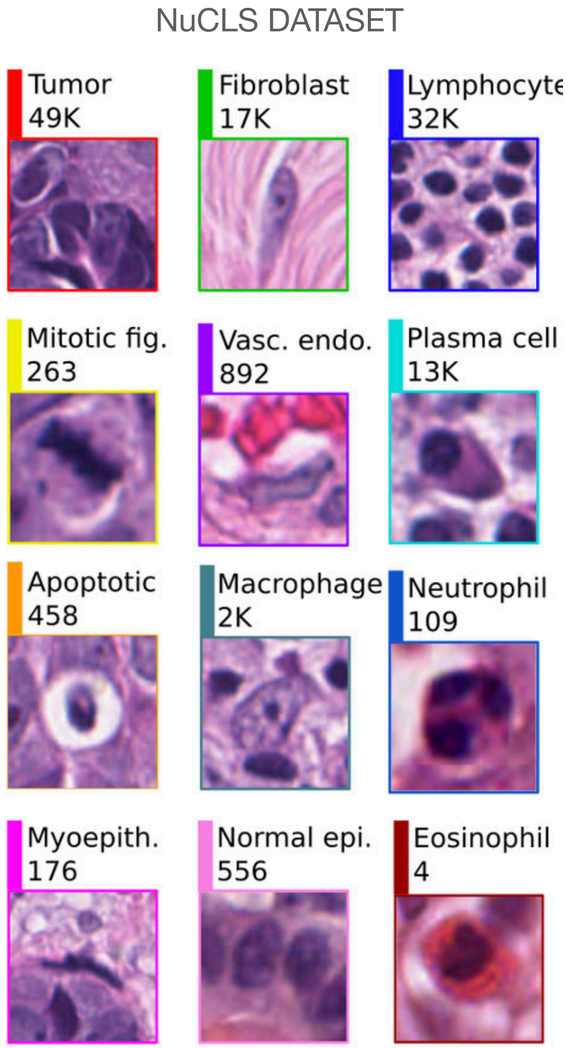
Output = Image (usually RGB)

A Cat Sitting on the sofa ⟶

Multiple possible outputs

Why text to image?

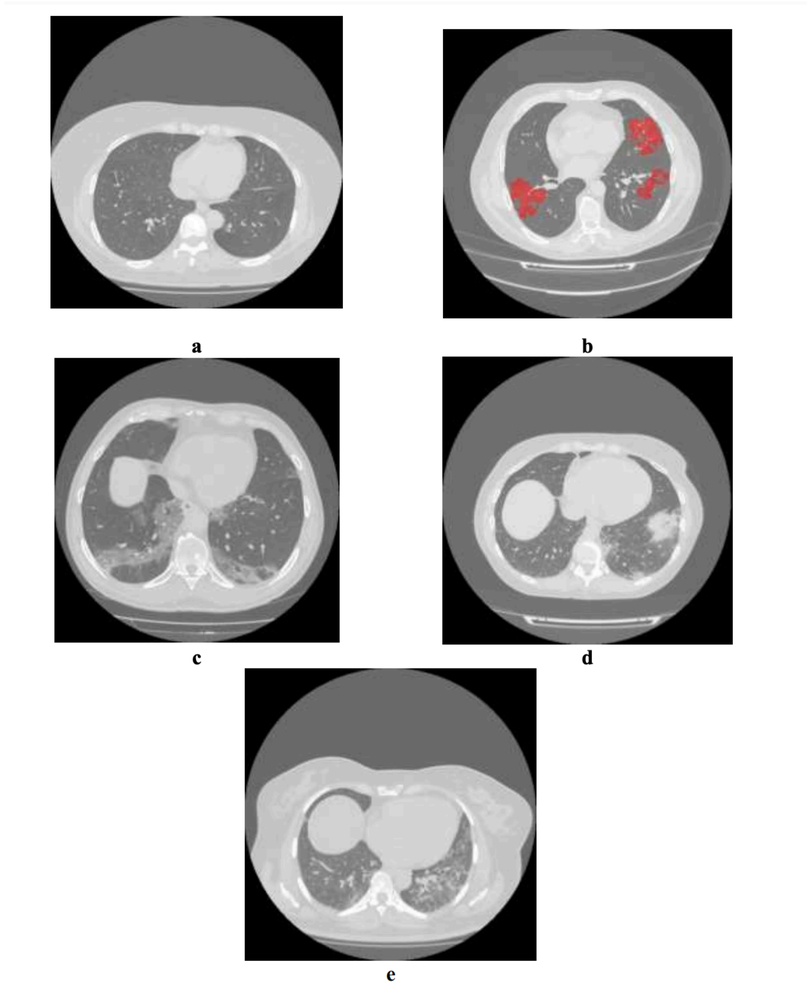
Applications of Perception.ai



The massive problem of class imbalance in Artificial Intelligence, with Perception.ai you are able to generate images of rare classes which otherwise occur infrequently or not at all. Example - Pathology slides for skin cancer. The ratio of the rarest class to most prominent class is 1:12250.

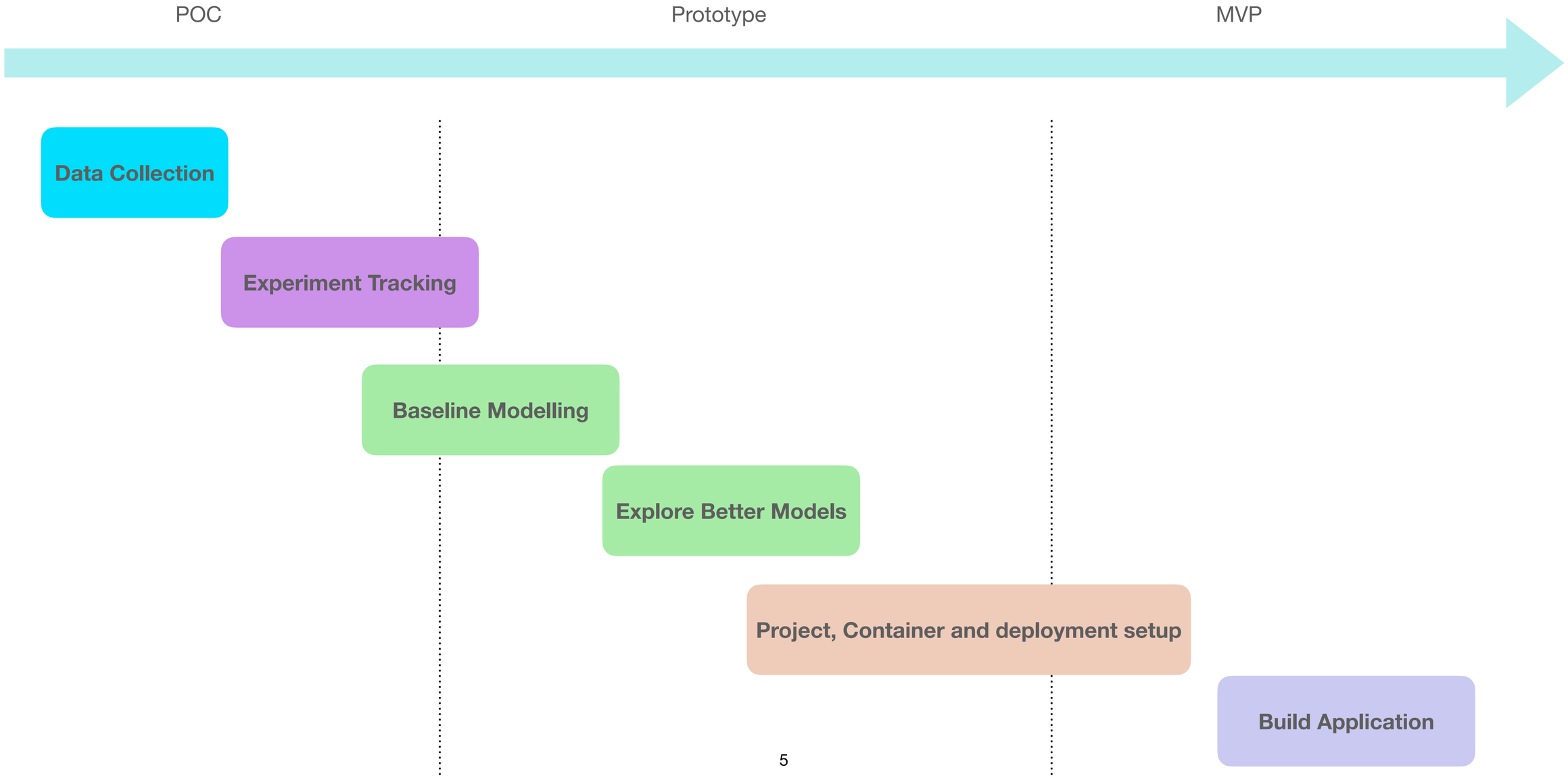
Why text to image?

Applications of Perception.ai



- Mosmed-1110 Dataset contains human lung tomography (CT) scan with covid19 related findings.
- Number of cases by Category:
  - CT-0 - 22.8%
  - CT-1 (Mild) - 61.6%
  - CT-2 (Moderate) -11.3%
  - CT-3 (Severe) - 4.1%
  - CT-4 (Critical) - 0.2%

Project Workflow

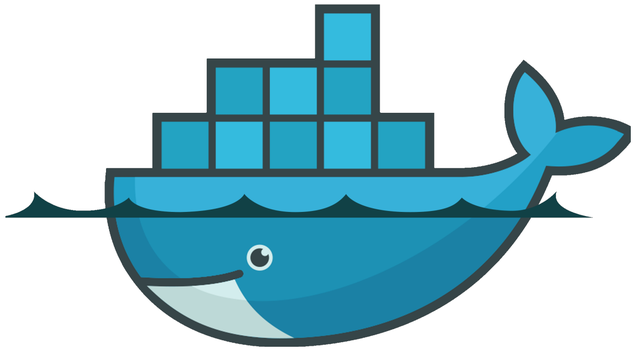


Technology Used

Tensorflow

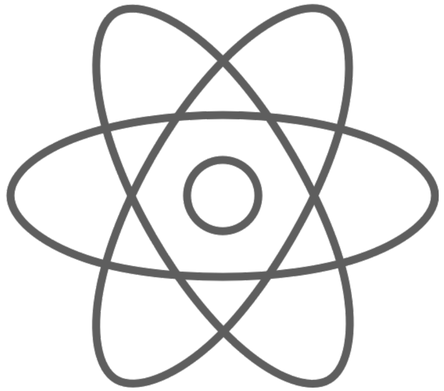
Pytorch

Ansible

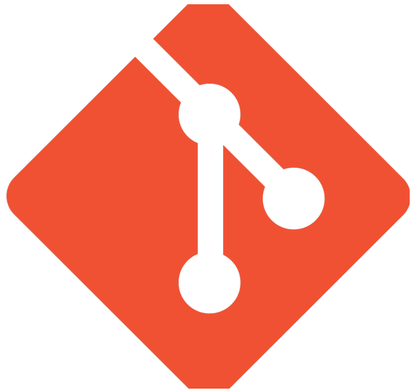


Docker

Google Cloud Platform



React



Git

Python

Data

CUB-200-2011 DataSet

- Number of Images - 11,788
- Number of Categories - 200
- One Bounding Box per Image for object detection
- Ten captions for each image.
- Pre trained sentence embedding using skip thought model.

Class ID: Laysan Albatross

Captions

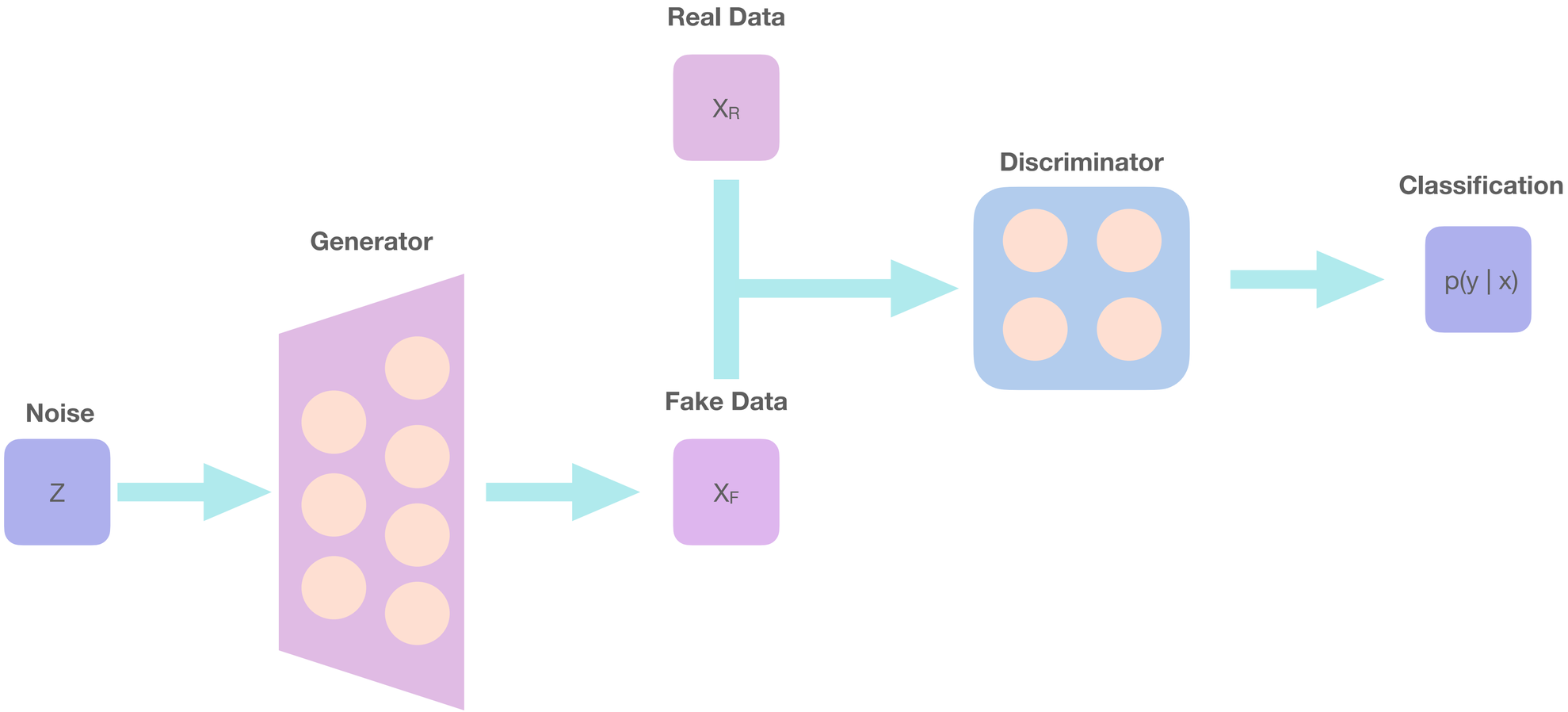
1. A white bird with black tipped wings and a long grey beak.

2. this white bird has black along the ends of its wings and a pale, long beak.

Modelling

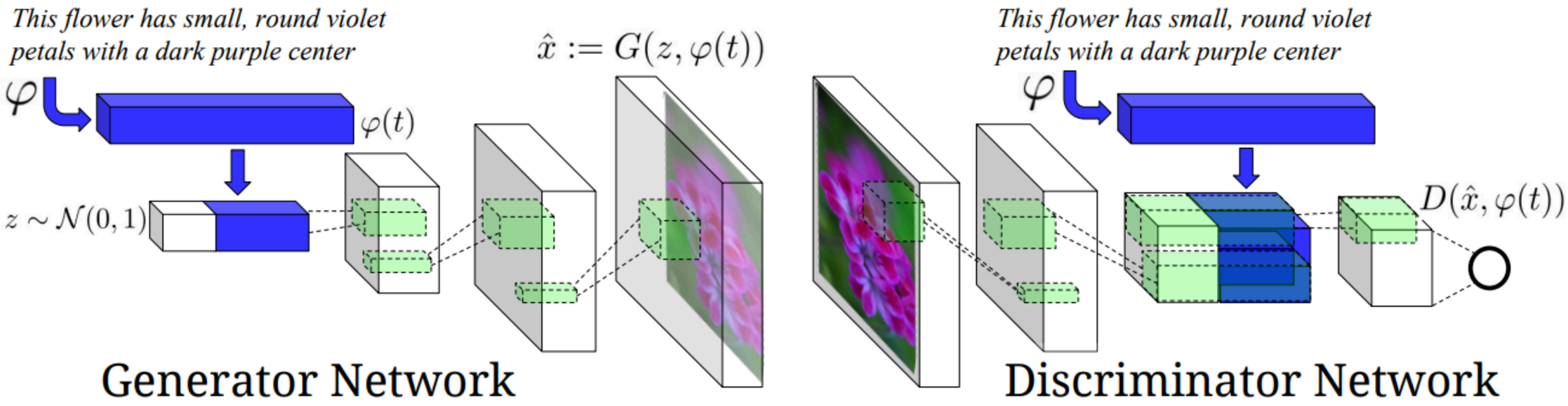
What are GANs?

A high level overview



Related Work

GAN-INT-CLS: RNN encoder with GAN decoder (2016)



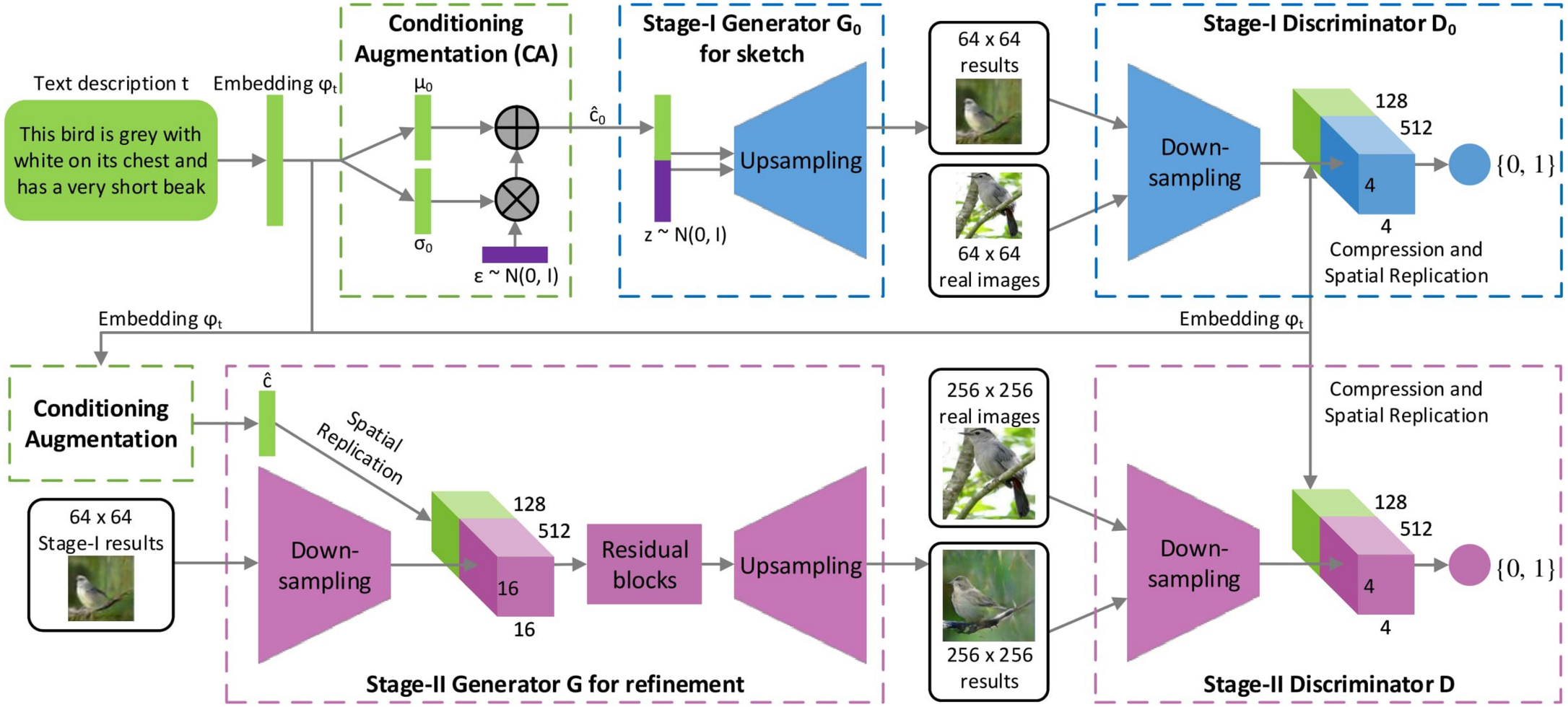
Related Work

GAN-INT-CLS: RNN encoder with GAN decoder (2016)

- First paper to propose the Idea of text to image using generative adversarial modelling.
- Simple model that takes in text embedding and uses a RNN encoder and generative model as decoder.
- The discriminator differentiates between the fake image and real image.
- Problem - the output images are not crisp and are not a good substitute of real images.

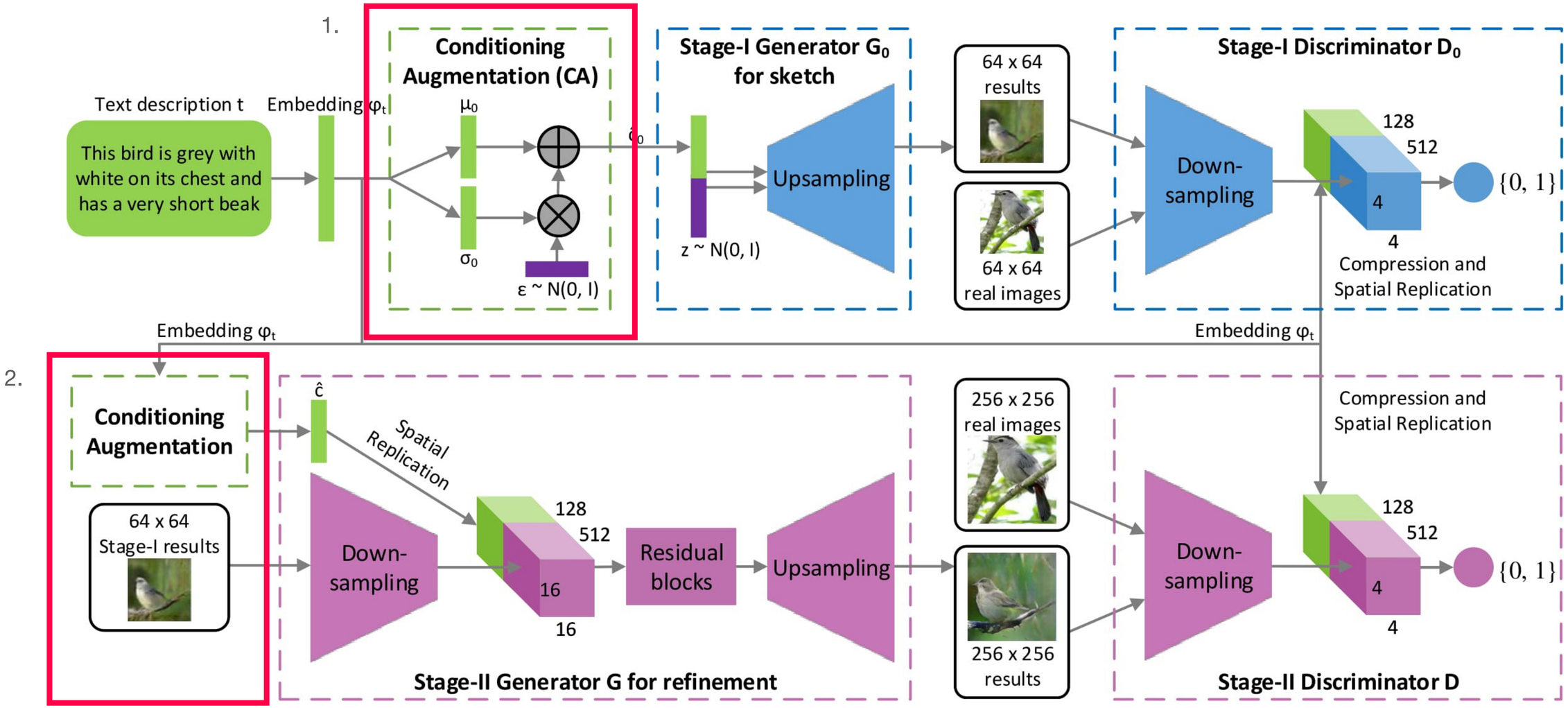
Baseline Model

StackGAN (2016)



Baseline Model

StackGAN (2016)



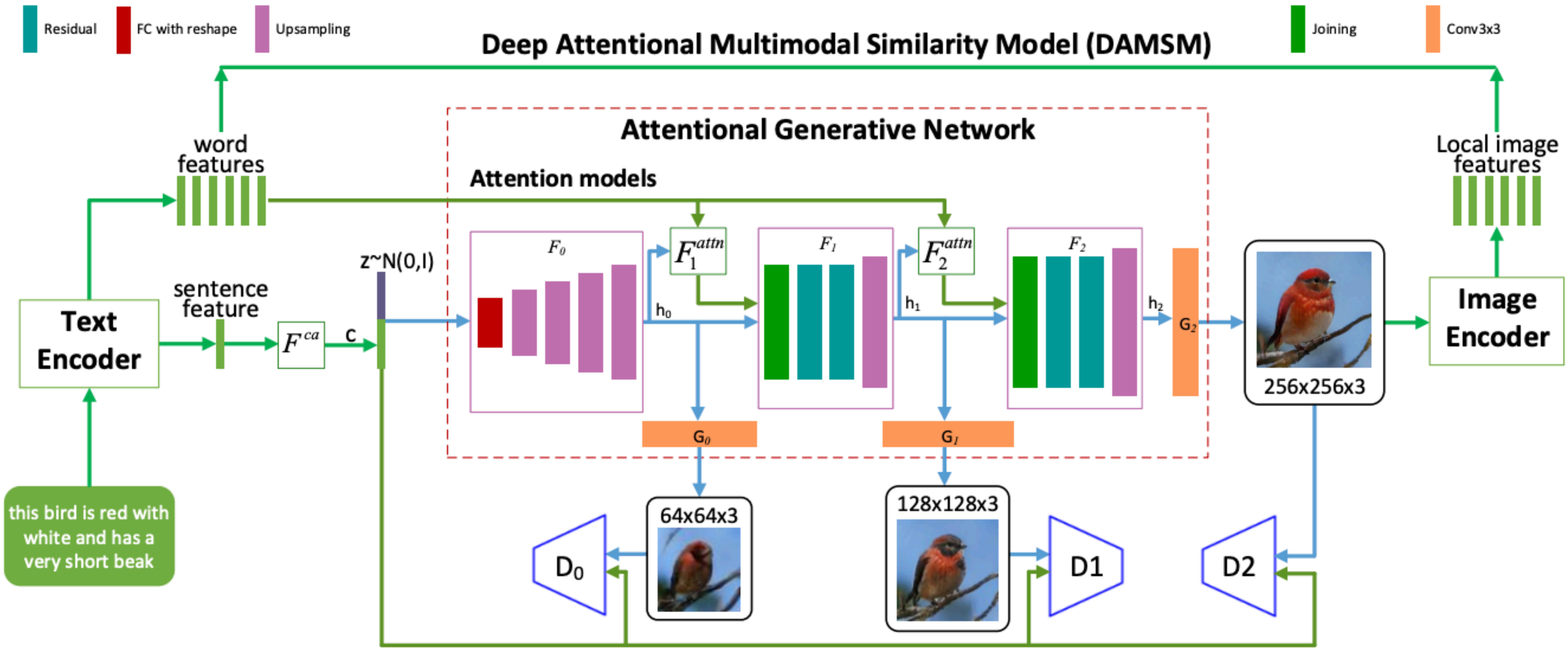
Baseline Model

StackGan

- Two important ideas
  - Conditional Augmentation Block which samples latent variables from a distribution, it makes the generator more robust in capturing various objects and poses and at the same time increases randomness to the network.
  - 2 Generative models stacked on top of each other to give high resolution image
  - Problem
    - Only a single sentence embedding is used as an input therefore there are no word level association between the sentence and the image.

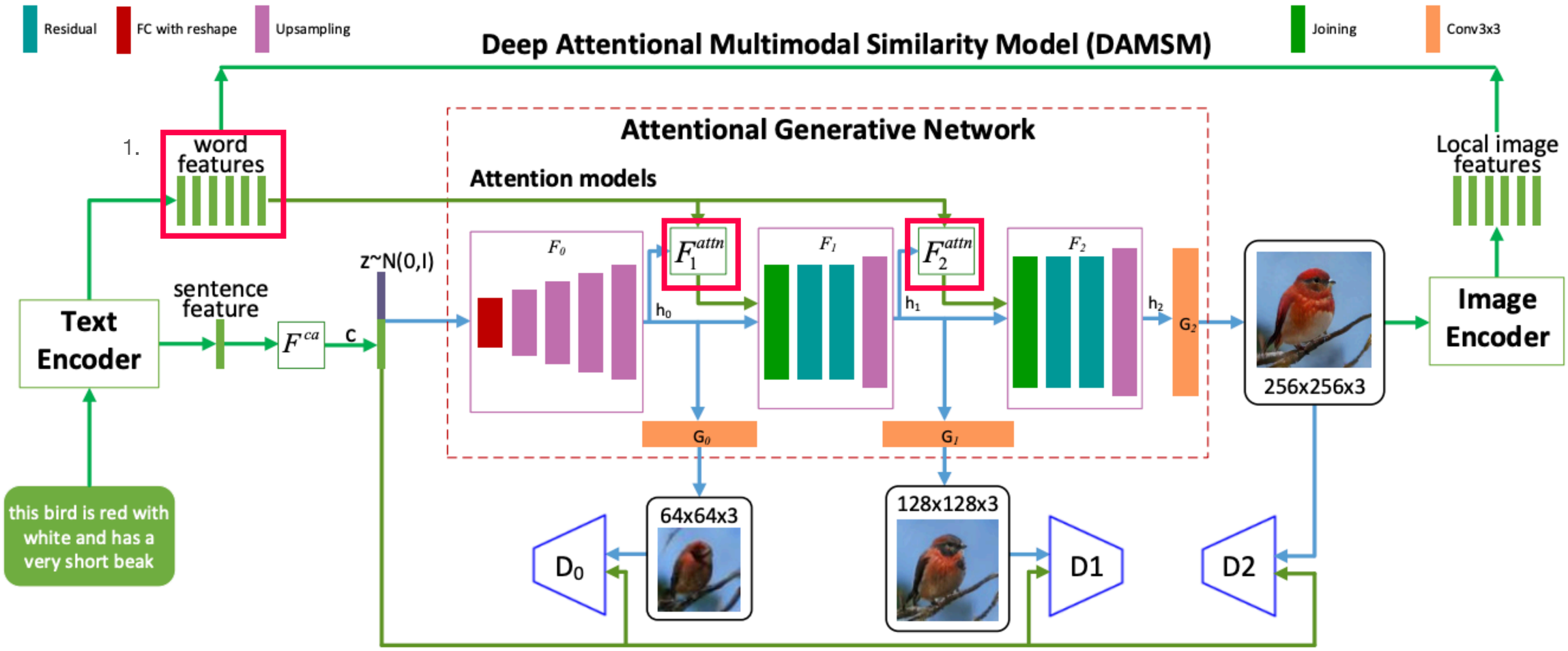
Improved Model

AttnGAN (2018)



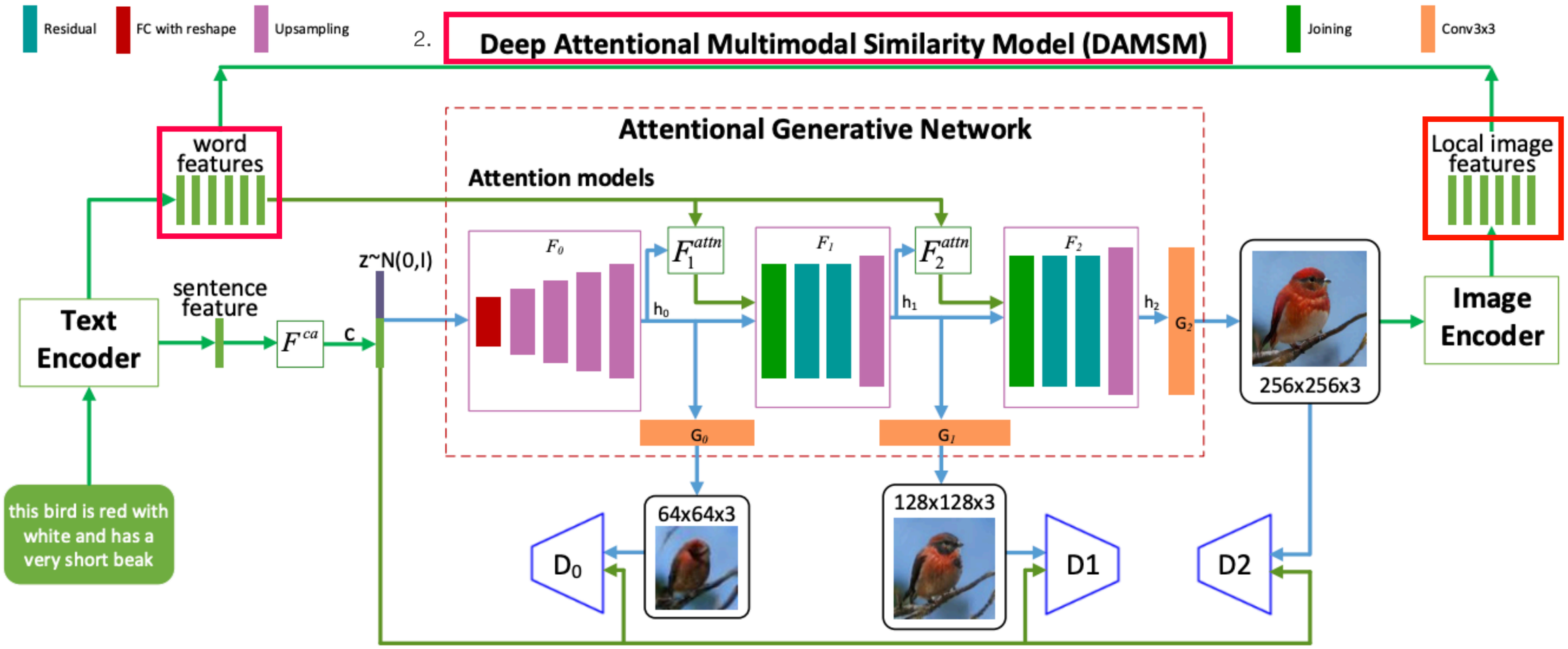
Improved Model

AttnGAN (2018)



Related Work

3. AttnGAN (2018)

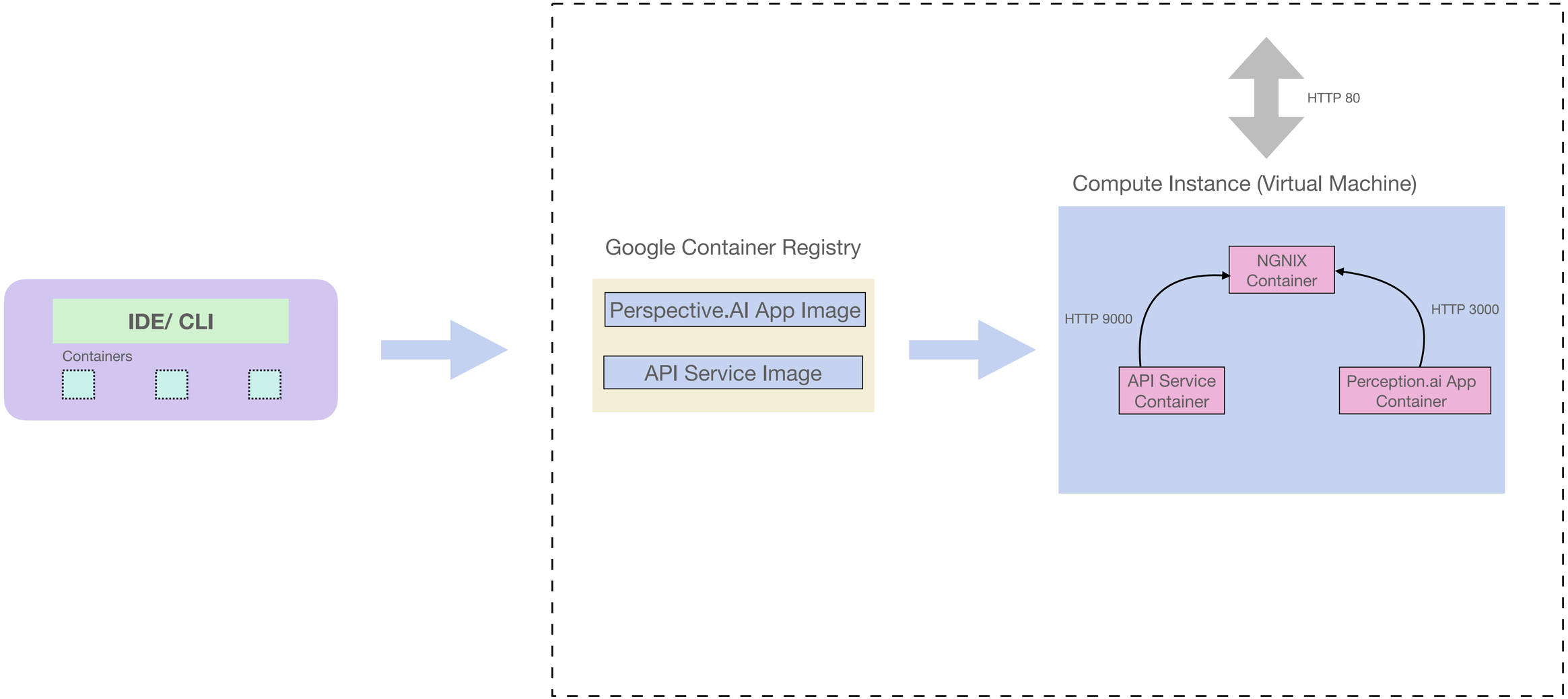


Improved Model
AttnGAN

- Two important ideas
  - Word level features are used to train the model along with sentence embedding
  - Final image is passed through an encoder and DAMSM loss is calculated, which helps us make relations between parts of the image and words in the sentence.

Project, Container & Deployment

Application Setup



Demo

Future Work

- Creating an API service where one can upload an image and caption data to train the Generative model remotely.
- We will like to extend this idea of generating images to generating videos using GANs. Such architectures can be useful in forecasting applications for weather prediction, autonomous driving etc.
- Finally, building generative models of the world around us is considered as one way to measure our understanding of physical common sense and predictive intelligence.