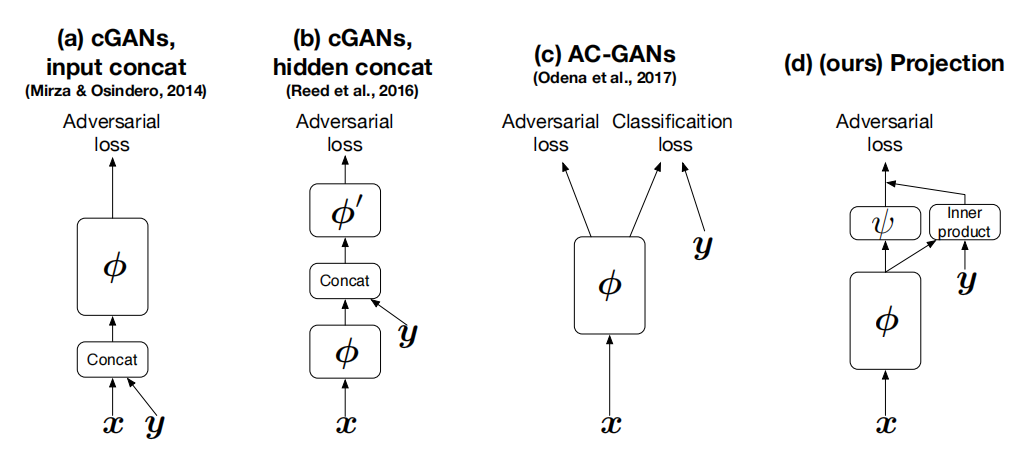
3. Method

3. x. Generative Adversarial Network (GAN)

GAN是一种通过学习潜在数据分布进而生成新的数据的模型。GAN的核心思想就是通过两个神经网络——生成器Generator和判别器Discriminator以相互对抗的方式进行训练。生成器接受一个随机噪声向量，并试图生成与真实数据分布类似的合成数据以欺骗判别器。

Conditional GANs (cGAN) 是一种拓展的GAN，在生成器和判别器中加入了额外的条件信息，使得生成结果更加可控，可以根据特定条件生成数据。

与传统的cGAN简单地将条件信息连接到判别器输入不同，cGAN with Projection Discriminator使用一种更为有效的方式结合输入数据和条件信息，来提升判别器的表现。



cGANs的判别器模型

cGAN with Projection Discriminator通过投影的方式将条件信息融入到鉴别器的决策过程中，具体而言，鉴别器计算样本的特征表示与条件信息的内积，并将这一结果加入到鉴别器的决策过程中，进而提高判别器的判别能力。

我们尝试使用cGAN with Projection Discriminator来实现有限数据下的带标签数据生成，并将生成的图像加入到输入的训练数据中，进而实现数据增广。

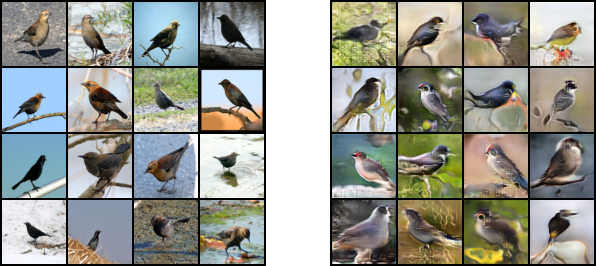
4. Experiment

4.x GAN

在使用cGAN with projection discriminator生成指定标签的图像时，我们使用了基于ResNet的生成器和判别器，并在判别器上应用了projection，将条件信息投影到一个嵌入空间，并将这个嵌入与图像特征进行内积操作，以更好地结合条件信息和图像特征。对于损失函数，我们采用了基于标准GAN损失的hinge版本。在实验中，我们使用了Adam优化器，超参数设置为α = 2e-4，β1 = 0，β2 = 0.9。每次更新生成器，我们都更新鉴别器5次。

我们在一个小数据集CUB\_200\_2011上进行测试。这个数据集中包含了200种鸟类图像，每种类别只有大约50-60张图片。我们将输入图像压缩成128\*128，并输出同样大小的生成图像。

在实验过程中，我们对模型进行了5000轮迭代。在迭代将近结束时，生成器和判别器的损失已经趋于稳定，我们大致推测此时模型已经接近收敛。



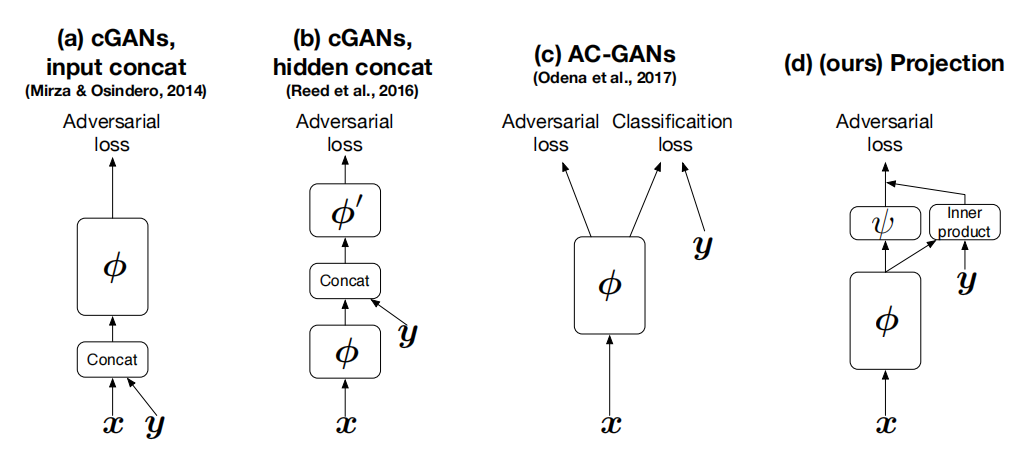
Comparison between Real Images (left) and Generated Images (Right) for class 11

图x是在类别11（Rusty\_Blackbird）上真实图像和生成图像的对比。可以发现，生成的图像在视觉上大体轮廓较为准确，但在细节上仍有欠缺。并且由于数据集的规模较小，每个类别的数据较少，导致最终生成模型在类间表现较差，不同类别的生成图像没有很好地呈现出不同类别的特征。

TODO 将生成图像加入到细粒度分类模型的训练过程中

GANs are models that generate new data by learning the latent distribution of existing data. The core idea of GANs involves training two neural networks — the Generator and the Discriminator — in an adversarial manner. The Generator takes a random noise vector and attempts to generate synthetic data similar to the real data distribution to deceive the Discriminator.

Conditional GANs (cGANs) are an extension of GANs that incorporate additional conditional information into both the Generator and the Discriminator, making the generated results more controllable and allowing for data generation based on specific conditions. Unlike traditional cGANs that simply concatenate the conditional information to the Discriminator's input, cGANs with Projection Discriminator use a more effective method to integrate input data and conditional information to enhance the Discriminator's performance.



cGANs的判别器模型

cGANs with Projection Discriminator integrate conditional information into the Discriminator's decision-making process through projection. Specifically, the Discriminator computes the inner product of the sample's feature representation and the conditional information, adding this result to the Discriminator's decision-making process to improve its discriminative ability.

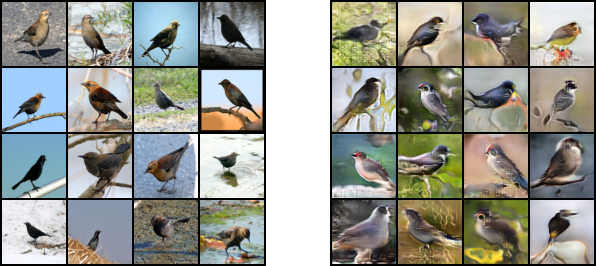
We attempt to use cGANs with Projection Discriminator to generate labeled data under limited data conditions and incorporate the generated images into the input training data, thereby achieving data augmentation.

Ex

When using cGAN with a projection discriminator to generate images with specified labels, we employed a Generator and Discriminator based on ResNet and applied projection on the Discriminator. This projection mapped the conditional information to an embedding space and computed the inner product of this embedding with the image features to better combine conditional information and image features. For the loss function, we utilized a hinge version based on the standard GAN loss. In our experiments, we used the Adam optimizer with hyperparameters set to α = 2e-4, β1 = 0, β2 = 0.9. We updated the Discriminator five times for every Generator update.

We tested our approach on a small dataset called CUB\_200\_2011, which contains 200 bird species images with approximately 50-60 images per category. We resized input images to 128\*128 and generated images of the same size.

During experimentation, we conducted 5000 iterations. Towards the end of these iterations, the losses of the Generator and Discriminator had stabilized, indicating that the model was nearing convergence.



Comparison between Real Images (left) and Generated Images (Right) for class 11

Figure x shows a comparison between real and generated images for category 11 (Rusty\_Blackbird). It can be observed that the generated images roughly capture the overall outlines accurately but lack in finer details. Additionally, due to the small scale of the dataset and limited data per category, the final generated model performed poorly across categories, failing to adequately represent the distinct characteristics of different categories in the generated images.