Measuring image data augmentation for generative models

Abstract

目标：

着眼于特征，定义与量化评价数据增强有效性，量化评价GAN对样本特征的扩征能力。
（How to measure feature expansion for image data augmentation？）

方法：

1.探讨能够度量图片语义的空间需要哪些性质？提出在特征空间内有效评价语义扩充的条件，揭示GAN对样本特征扩充的物理本质（可解释性、物理原理），指明了隐空间，作为一种特殊的特征空间，具有语义表征的能力（摄影几何vs分布拟合，传统增强方法与GAN增强方法的不用）

2.分析验证GAN-like方法隐空间的性质与语义扩充评价条件，提出对GAN-like网络建立公共隐空间的方法
（验证公共隐空间的性质），实现对不同模型特征扩征的比较

3.解决了OT距离的存在性，提出评价指标的解析解，确立公共隐空间中的特征扩充评价指标（测度metrics），量化评价不同算法扩征有效性

结果与结论：

对不同类型的GAN网络测试了评价模型与评价指标，分析性能差异，并与传统的评价模型做对比。探讨了能够度量语义信息的特征空间需要满足的性质，有效评价语义扩充的条件。在公共特征空间中，提出了着眼于语义信息评价数据增强效果的评价指标。指导在数据增强问题上GAN的评价与选择。

OUTLINE:

1 Introduction

1.2 Central question

objective of data augmentation:

solve insufficient data, solve class imbalance, feature augmentation

How to measure feature expansion for image data augmentation?

-- What are the criterions on feature space for evaluation of image data augmentation，and our solution.

问题的逻辑，problem formulation.观点-原理-例子

论文的立意：争论的点 悬而未决的问题 参考顾老师提出的方法

正文：

1.Introduction

Deep Learning have made great progress in NLP, CV, SLAM and have brought enormous changes to our lives. While DL has proven its potential and ability in accomplishing computer vision tasks, it is generally accepted that the performance of Deep Learning Models relies heavily on the dataset. In industry scenario, the data is usually expensive and laborious to collect due to data privacy, require of expertise. A limited and imbalanced dataset becomes a major obstacle in the implementation of AI.

Data Augmentation, aimed at extracting more information from the original data to enhance the training dataset have shown great results in tackling both limited dataset and classed-imbalance and is gradually becoming the key to the industrialization of Artificial Intelligence. Nevertheless, traditional data augmentation methods, encompassing geometric or color-space transformation, have reached a bottleneck in exporting and enriching the semantic feature of the data and spray a light on DL-based augmentation methods, especially Generative Models.

Generative Models have been widely and successful adapted in Automatic Driving, Medical Research or Computer Vision fields. While GANs have shown impressive ability in synthesizing exquisite fake samples that confuses human expert, they also suffers Model Collapse and hard to reach the Nash-equilibrium in Training. Generating exquisite fake images marks a successful stage of GAN models, and is usually evaluated by FID, FIS or directly by human eyes, yet the enhancement that GAN has brought to dataset(partially, how GANs conquer the Model Collapse) has merely been discussed. Recent GAN models express great ability to expand the feature of the original dataset, such as Info GAN (attribute editing) or styleGAN (feature synthesis). Well-trained GAN models can not only solve limited-data or class-imbalance, but also enrich the semantic feature of the original dataset.

Address the problem： 解决什么痛点?

The impressive performance of generative models expand our imagination on the use of machine creativity in data augmentation , nevertheless has generative models really augmented the data? Moreover, has the "creativity" of GAN really expand the semantic in the dataset? As a first investigation the this large problem, this paper tries to propose a evaluation model that measures the augmentation ability (concerning data sufficiency, class balance and feature expansion) for generative models. First, we propose the criterions on feature space based on the purpose of data augmentation and the nature of GAN-based augmentation method (chapter 2). Secondly, we designed a semantic feature space satisfying the proposed criterions, also a OT-based measure is proposed in the semantic feature space compare the original and augmented distribution (chapitre3). Finally, we proposed a data augmentation evaluation method for GAN models, different GAN algorithms were tested to guide the selection and design of GAN in image data augmentation.(chapter 3 and 4) We found that stylegan and ... lead GAN-based augmentation method in .... .

The principle contributions of the paper are:

- A semantic feature space is proposed to measure the semantic feature expansion in data augmentation, the method only depends on the original data set, making the evaluation method both relevant and generalized.
- A OT-based feature enhancement measure, LASFAS, is designed in the semantic feature space, which has more reliability on the evaluation of feature augmentation.
- Different generative models as well as sampling methods are evaluated to guide the selection and use of generative models for data augmentation.

OUTLINE:

2 Feature space for evaluation of image data augmentation(如果2.1只做review没有自己的东西，做连贯的一章)

Feature space of image data（删？）

- Latent space of gan
  - 隐空间，满足的性质1，性质2，性质3 -> 通过前人对隐空间性质的研究验证。
  - 隐空间具有语义表征的能力：属性编辑等。
  - 揭示GAN对样本特征扩充的物理本质：**流形学习角度解释**GANs如何补全特征+小实验。
- Network-based feature space
- **What are the problems?** latent space oddity: difficult to design metrics, feature space: hard to explain, some prior assumptions are not varified: vgg-network based FID and IS score.
- However some naive idea for feature measurement shall be preserved
  - linear separable of latent space
  - linearity of latent space

Analysis: Data augmentation（G）与 F 之间的逻辑关系

objective of data augmentation:

solve insufficient data, solve class imbalance, feature augmentation（controversial）

The nature of feature-based data augmentation

- high dimensional data -> a distribution on low dimensional non linear manifold
- genrative model, transfor uniform distribution to the original data distribution -> oversampling the learned manifold with interpolation and expolation.

How to evaluate

- measuring the completion of these tasks:
  - insuficient data：efficient interpolation-linearity,
  - solve class imbalance: class separability, K-means
  - feature augmentation: between class convexity
- how to evaluate interpolation or expolations on the manifold? (no inclusion of tradidtional method that requires prior knowleadge)
  - require a feature space that flatten the manifold. (difficulty: manifold too abstract,
  - interpolation: constrained in a single class. For each class, linearity of different semantic features
  - expolation:each class forms a convex set in the $\mathbb{F}$ space, when new feature appears, it falls out of the current zone

Conclusion: Criterions on feature space

- Criterions on feature space for evaluation of image data augmentation: how to evaluation data augmentation based on feature space? 能够度量图片语义的空间需要哪些性质?
- we propose a new feature space to evaluate feature expansion of images data: $\mathbb{F}$ space

正文：

2 Related Works

Data augmentation

Data augmentation techniques were first developed to solve overfitting in Deep Learning. A model overfits the training data when it performs well on previously seen data but poorly on unseen data. The two main cause of overfitting are sucifficient data and class-imbalance. Insufficient data is a prevalent challenge in Computer Vision. It is a generally accepted that larger datasets result in better Deep Learning models.However, assembling enormous datasets can be a very daunting task due to the manual effort of collecting and labeling data. for example in Medical Analysis, many of the images studied are derived from computerized tomography (CT) and magnetic resonance imaging (MRI) scans, both of which are expensive and labor-intensive to collect. Class-imbalance refers to the abnormal ratio of majority and minority samples in a data set. The model trained on a dataset with imbalanced classes usually have the tendency to "sacrifice" minor samples to ensure accuracy on majority classes. As a consequence, the model usually fails to perform on minor classes. Moreover, improving the generalization ability of Deep models is one of the most difficult challenges today. Tasks like domain adaptation and generation require the model to extract as many useful information as possible for the learning of a unseen domain.

Traditional data augmentation techniques based on geometrics or color transformation, such as rotation or translation can artificially inflate the dataset and partially solve insufficient data or class imbalance.(citation) However the prior knowledge set by human constrains its ability in enriching sematic features. The generative-based augmentation method, on the other hand has shown great potential in semantic feature generation and controlling. ... et al. proposed a feature synthesis network styleGAN and has greatly enrich the facial semantic features of the original dataset. ... et al. proposed a attribute editing method based on GAN models that enable feature controlling of the generated samples.

Thus we define the three main tasks for data augmentation as:

- Data-sufficiency
- Class-balance
- Semantic feature expansion

To measure the augmentation ability of a GAN model, we will then measure its ability on the three sub tasks of data augmentation. Before heading down to the specific measuring method, we want to discuss the nature of generative-based augmentation method, and how and why GAN-based models have great potential on three main tasks of data augmentation.

Nature of GAN-based augmentation method

The real data satisfies the manifold distribution hypothesis: their distribution is close to a low dimensional manifold in the high dimensional image space.

Ideally, if given an embedding map $f : \chi \rightarrow \Omega$ and a dense dataset X sampled from a distribution vgt supported on χ, the purpose of the generation model is to generate new samples following the distribution of vgt and locating on the manifold χ. For the AE-OT model, it only requires that the reconstructed images should be similar to the real ones under L2 distance. As a result, the support of the generated image distribution may only fit the real manifold χ well near the given samples. For GAN model, on one hand, the feature loss and content loss require that the reconstructed manifold should approach to the real manifold χ on the given samples; on the other hand, the discriminator is used to regularize the fitting performance of the generated manifold on both the given samples and new generated samples should fit the real manifold well. Therefore, the generated manifold by the GAN model fits the real manifold χ far more better than the AE models. In conclusion, first of all, with the

help of the discriminator, GAN models fit better the real data manifolds. Secondly, when generating new samples, GAN models transform the prior distribution in latent space to distribution on the learned data manifold.

Thus the main method for GAN-based image data augmentation is to oversampling on the learned data manifold. The tow mainstream manipulation for oversampling are interpolation and extrapolation of semantic features. interpolation refers to the sampling method that interpolates between two different points on manifold distribution, which ideally bring a linear changes in the picture. Extrapolation refers to operation that fills up the flaw on the manifold, these flaw may be caused by class imbalance or limited data of the original data set. Such operation results in the synthesis or enhancement of features. The styling-mixing operation of styleGAN is a good example of designed feature extrapolation technique.

Criterion on semantic feature space

Knowing the goal for data augmentation as well as the oversampling nature of GAN-based method, we are close to define the question: based on the oversampling nature, how do we measure the performance of GAN models on Data-sufficiency, Class-balance and Semantic feature expansion. It is naturally to consider to compare the original and augmented data distribution on the data manifold, but the question is the manifold is highly abstract theoretical concept and hard to find. Therefore, we propose on alternative to designed a feature space that fits certain criterions targeting the three major tasks in data augmentation.

- In class linearity:

  GAN-based model usually tackle data insufficiency by randomly sampling in latent space. However the sampled data point might not be efficient for data augmentation, for example when most images are sampled around a single point in data manifold, also known as model collapse in GAN training.

- Between class separability:

  In order to clearly measure the data point in each class for class balance, we expect the distance between centers of different classes to be as far as possible. Moreover, we also expect the two different class set to be separated by a hyperplane.

- Convexity:

  When features are enhanced or synthesized, which we believe they should be consider different from each source class, we expected these data point to fall out of the root feature zone. In other words, each class shall form a convex set that feature-preserving operation, such as in class interpolation, is also class preserving in the feature space. Feature augmenting operations such as style mixing or feature enhancement will generate out-class points.

OUTLINE:

3 Toward $\mathbb{F}$ space

Based on the criterions proposed in section 2, we attempt to, in this section, find the $\mathbb{F}$ space through training a network that maps images to $\mathbb{F}$.

3.1 criterions on $\mathbb{F}$ space (feature space design)

measuring the completion of these tasks:

- insufficient data: efficient interpolation-linearity,
- solve class imbalance: class separability, K-means
- feature augmentation: between class convexity

3.2 OT-based metrics in feature space $\mathbb{F}$ (metric design)

measuring the completion of these tasks:

- insuficient data：efficient interpolation-linearity ---- K-means
- solve class imbalance：class separability ----
- feature augemention：between class convexity ----OT comparison?
- **解决了OT距离的存在性**（附件）
- 提出评价指标的解析解（附件）
- 确立公共隐空间中的特征扩充评价指标（测度metrics)--LASFAS
- 量化评价不同算法扩征有效性

(3.3) feature score in LASFAS

将验证评价指标与FID,IS等指数的关联性，例如 SIC = FID + IS + FS，其中FS代表了对语义信息的评价

3.3 Method to evaluate generative models in GANs

过渡段（生成模型的评价）：流形能够被生成模型捕获或者利用

提出对GAN-like网络建立公共隐空间的方法

- 对于两个预训练的生成器网络A，B
- 寻找映射f, 以方法一中提所出的性质为目标将A，B映射到**公共中间空间**W，W是一个特征空间。
- 在**公共中间空间**W中选择比较分布的距离。

OUTLINE:

4 Experiment

To test and evaluate the performance of LASFAS, we compared it with traditional evaluation metrics (FID and IS) on several data sets with different GANs. Moreover, we have explained its relevance with FID and IS using... theorem.

4.1 Minist（寻找更好的数据集）

(验证语义合理性--文献调查) visual salience

通过可视化的方式验证 $\mathbb{F}$ 的合理性

- linearity
- convexity of class set

4.2 CelebA dataset

结果

从实验上证明打分的有效性：经过不同生成网络增强之后的数据集在 sematic feature space 中的分布投影。

-
GAN 方法	隐空间的映射	对隐空间向量的操作：feature	
style-gan	LSFAS, IS，FID	LSFAS, IS，FID	
wgan	LSFAS, IS，FID	LSFAS, IS，FID	
...	LSFAS, IS，FID	LSFAS, IS，FID	

Feature score in LASFAS

将验证评价指标与FID,IS等指数的关联性，例如 SIC = FID + IS + FS，其中FS代表了对语义信息的评价

5 Results

- 提出设计了度量语义的特征空间 $\mathbb{F}$,在minist上验证了 $\mathbb{F}$ 对语义度量的合理性
- 对不同类型的GAN网络测试了评价模型与评价指标，分析性能差异，并与传统的评价模型做对比

6 Conclusion

- 探讨了能够度量语义信息的特征空间需要满足的性质，有效评价语义扩充的条件。在公共特征空间中，提出了着眼于语义信息评价数据增强效果的评价指标。feature space 只依赖于原数据集，不依赖于生成的模型与方法。相较于FID 和 IS 更加具有针对性，但又不失泛用性。
- 能够通过GAN族生成的数据直接度量生成方法data-augementation的有效性(feature expansion)。指导在数据增强问题上GAN的评价与选择。

Contribution:

- 在语义空间中基于最优传输理论设计了特征增强测度——LASFAS， 能够合理的评价图片数据特征增强效果。
- 基于流形学习假设和生成器原理提出了度量语义信息的特征空间 $\mathbb{F}$, $\mathbb{F}$ 仅依赖原有数据集而不依赖于生成模型，使得评价方法既有针对性也不失泛用性。
- 对不同生成模型以及采样方法进行了评价，指导生成模型在数据增强上的选择与使用。

- feature space 只依赖于原数据集，不依赖于生成的模型与方法。这使得我们的方法相较于FID 和 IS 更加具有针对性，但又不失泛用性。
- 通过GAN族生成的数据直接度量生成方法data-augementation的有效性(feature expansion)。
- 探讨了能够度量语义信息的特征空间需要满足的性质，有效评价语义扩充的条件。在公共特征空间中，提出了着眼于语义信息评价数据增强效果的评价指标。
- 指导在数据增强问题上GAN的评价与选择。

为什么选择feature space + metric 的方法

- feature space 只依赖于原数据集，不依赖于生成的模型与方法。这使得我们的方法相较于FID 和 IS 更加具有针对性，但又不失泛用性。
- 通过GAN族生成的数据直接度量生成方法data-augementation的有效性(feature expansion)。

重点参考哪几篇文献——Reviewer

- nature - machine intelligence -- 调查文献
- PAMI -
- **重点文献与重点审稿人**

重点参考哪几篇文献——Reviewer

- nature - machine intelligence -- 调查文献
- PAMI -
- **重点文献与重点审稿人**