

Машинное обучение (Machine Learning)

Деформирующие автоэнкодеры (Deforming autoencoders - DAE)

Уткин Л.В.



Порождающие модели

- 1 Вариационные Автокодеры - Variational Autoencoders – VAE
- 2 Генеративные состязательные нейросети - Generative Adversarial Networks - GAN
- 3 Деформирующие Автокодеры - Deforming autoencoders - DAE

Классические автокодеры

- Могут обучаться для генерации компактных представлений
- Хорошо восстанавливают исходные данные
- **Но** фундаментальной проблемой автоэнкодеров является то, что скрытое пространство, в котором они кодируют входные данные, может не быть непрерывным и не допускать гладкой интерполяции!

Вариационные Автокодеры (VAE)

- Могут решить эту проблему, так как их скрытое пространство является непрерывным и позволяет легко производить случайную выборку и интерполяцию
- **Но** управление глубокими нейронными сетями и, в особенности, глубокими автокодерами - сложная задача, ключевая особенность которой — строгий контроль процесса обучения

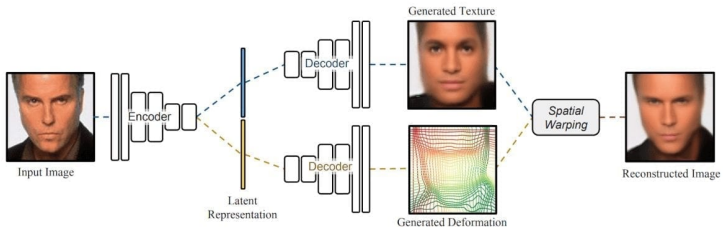
Деформирующий автокодер

- Это - породающая модель анализа изображений, которая выделяет признаки без дополнительных подсказок, предполагая создавать экземпляры объектов посредством деформации “шаблонного” объекта.
- Это означает, что вариативность объекта может быть разделена на уровни, связанные с пространственными трансформациями формы объекта.
- **Z. Shu et al. Deforming Autoencoders: Unsupervised Disentangling of Shape and Appearance // arXiv:1806.06503, Jun. 2018.**

Подход детальнее

- Деформирующий автокодер способен определять форму и внешний вид объекта как степени вариативности в изученном малоразмерном скрытом пространстве
- Архитектура состоит из
 - кодера, который кодирует входное изображение в два скрытых вектора (один – для формы, другой – для вида)
 - двух декодеров, принимающих векторы в качестве входных данных и выдающих сгенерированную текстуру и деформации
- Независимые декодеры необходимы для получения функций внешнего вида и деформации

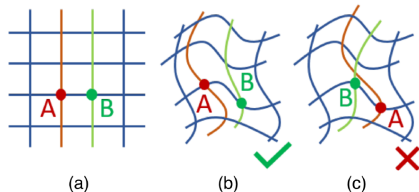
Структура (1)



Структура (2)

- Сгенерированная пространственная информация используется для деформации текстуры к наблюдаемым координатам изображения
- DAE может восстановить входное изображение и в то же время определить форму и вид объекта как различные особенности
- Вся нейросеть тренируется без помощника на основе лишь простых потерь восстановления изображения

Разрешенные деформации

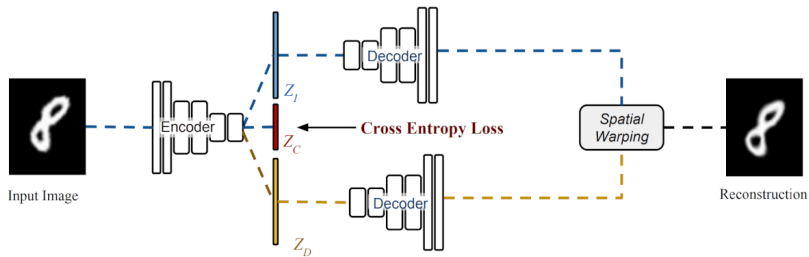


- Допускаются только локально согласованная деформация (b)
- Изменение относительных позиций пикселей (c) не допускается
- Для этого позволяем декодеру деформации предсказывать горизонтальные и вертикальные приращения деформации ($\nabla_x W$ и $\nabla_y W$)

DAE с заданной классификацией

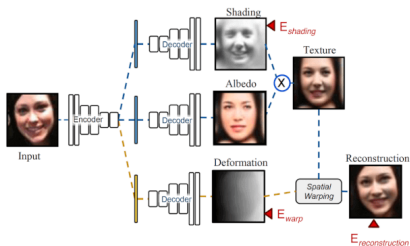
- **DAE с заданной классификацией** учатся восстанавливать изображение и одновременно определяют форму и вид факторов вариативности, соответствующие определенному классу
- Для реализации, вводят классифицирующую нейросеть, вход которой - третий скрытый вектор, используемый для кодирования класса. Это позволяет изучать смешанную модель, обусл-ную классом вх. изобр.
- Это улучшает эффективность и стабильность обучения, так как нейросеть учится разделять типы пространственной деформации, различные для каждого класса

DAE с заданной классификацией



Встроенный DAE

Для вычисления альбеда и теней на портретных изображениях



$$E_{\text{DAE}} = E_{\text{Reconstruction}} + E_{\text{Warp}}$$

- Reconstruction loss and warping loss

$$E_{\text{Reconstruction}} = \|\text{Output} - \text{Input}\|^2,$$

$$E_{\text{Warp}} = E_{\text{Smooth}} + E_{\text{BiasReduce}}$$

- E_{Smooth} штрафует быстро меняющиеся деформации, закодированные локальным полем деформации, измеряется как общая норма вариации горизонтальных и вертикальных деформаций
- $E_{\text{BiasReduce}}$ направлены на устранение любого систематического искажения, вносимого процессом подбора

MNIST

- DAE способен успешно выявлять форму и внешний вид объектов во время обучения
- DAE с заданной классификацией дают наилучшие результаты как при восстановлении, так и при изучении внешнего вида объектов



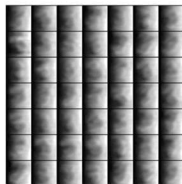
(a) input



(b) reconstruction



(c) decoded appearance



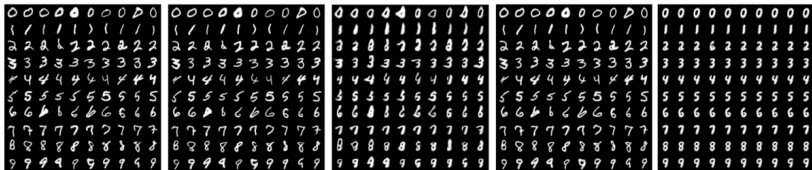
(d) decoded deformation

DAE в различных режимах (1)

- Наложение изображений без учителя.
- Изучение семантически важных множеств для формы и внешнего вида объектов.
- Внутренняя декомпозиция без учителя.
- Детектирование локализации без учителя.

DAE в различных режимах (2)

Восстановления MNIST с заданной классификацией



(a) input

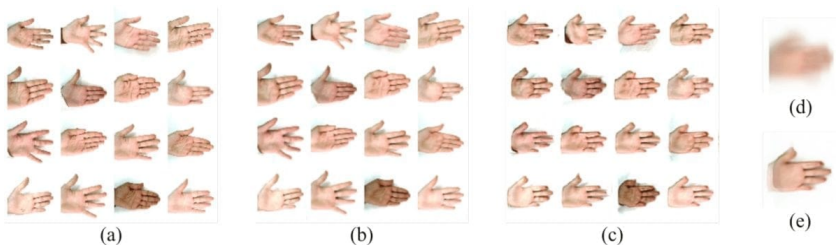
(b) Reconstruction
w/o classification

(c) Appearance
w/o classification

(d) Reconstruction
with classification

(e) Appearance
with classification

Наложение изображений ладоней без учителя



(a) – Входные изображения (b) – восстановленные изображения (c) – изображения текстур, деформированные с использованием среднего декодированной деформации (d) – среднее входное изображение (e) – средняя текстура

Интерполяция освещения с помощью встроенного DAE



И так...

- DAE - специфическая архитектура, способная выявлять определенные факторы вариативности – в данном случае это форма и внешний вид объектов.
- Результаты работы DAE показывают, что она способна успешно выявлять факторы вариативности посредством применения архитектуры автокодиров

Вопросы

?