МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

**Лабораторная работа №\_\_5\_\_**

по дисциплине«Разработка нейросетевых систем»

Тема: «Перенос обучения»

ИСПОЛНИТЕЛЬ: \_\_\_\_ \_\_

ФИО

группа ИУ5-25М \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

""\_\_\_\_\_\_2024 г.

ПРЕПОДАВАТЕЛЬ: \_\_\_\ \_\_\_

ФИО

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_"\_\_\_\_\_\_\_\_\_2024 г.

Москва - 2024

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Задание

Для набора данных с помощью автоэнкодера получить эмбеддинг изображений и его визуализировать.

Загрузить собственную аудиозапись и использовать автоэнкодер для удаления шума из аудиозаписи.

* Провести обучение модели по варианту.
* Изменить гиперпараметры обучения для понижения ошибки модели: количество эпох, размер батча, скорость обучения,
* Изменить количество слоев/нейронов, чтобы уменьшить и увеличить размерность скрытого пространства. Сравните полученные результаты
* Обучить модель автоэнкодера для удаления шума из аудио.
* Изменить гиперпараметры обучения для улучшения коэффициента детерминации: количество эпох, размер батча, скорость обучения
* Изменить количество каналов, чтобы уменьшить и увеличить размерность скрытого пространства. Сравнить полученные результаты

Отчет должен содержать: титульный лист, задание с вариантом, скриншоты и краткие пояснения по каждому этапу лабораторной работы, результаты эмбеддинга изображения, результаты удаления шума из аудиозаписи, итоговую таблицу со результатами для всех вариантов обучения

Классы согласно варианту из CIFAR100:

1. 39:
2. 70:
3. 35:

# Часть 1. Автоэнкодер для визуализации данных

Также как и в первых четырёх лабораторных я импортировал необходимые библиотеки, загрузили и распаковали набора данных CIFAR100 согласно варианту.

Была произведена аугментация с базовыми параметрами.

Затем для решения первой задачи была описанна модель полносвязного автоэнкодера с 6 скрытыми слоями (3 для энкодера и 3 декодера).

Функция потерь используется – MSE, а в качестве оптимизатора используется Adam.

Далее было проведено обучение с базовыми гипперпараметрами.

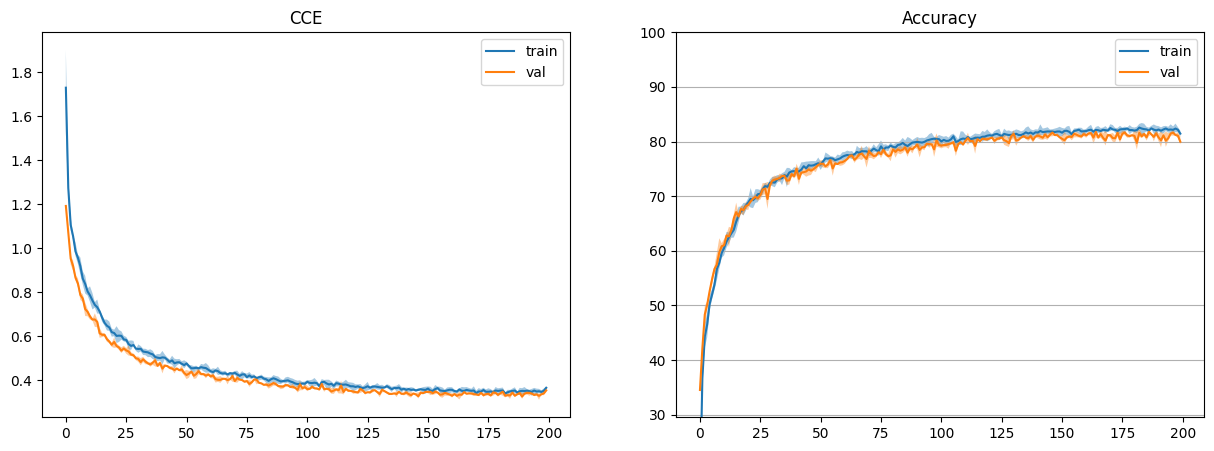
**Обучение модели по эпохам**

Рисунок 1 – График обучения НС базовой конфигурации

По рисунку 2 видно, что общие очертания картинок прослеживаются, но качество снимков можно попытаться улучшить.

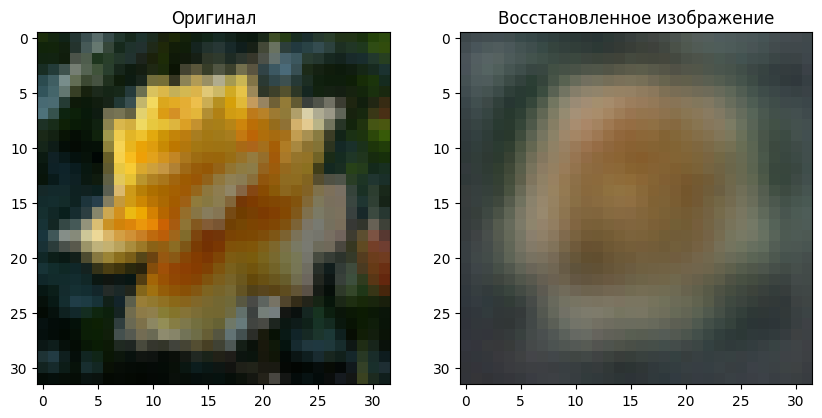


Рисунок 2 – Сравнение входной и восстановленной картинки

Был проведён ряд попыток улучшить результат, путём изменения гипперпараметров и также добавления дополнительного слоя. Характеристика проведённых опытов представлена в таблице 1.

В последнем опыте я попробовал добавить новый слой, увеличив размерность скрытого пространства. В результате обучения получил следующие графики.

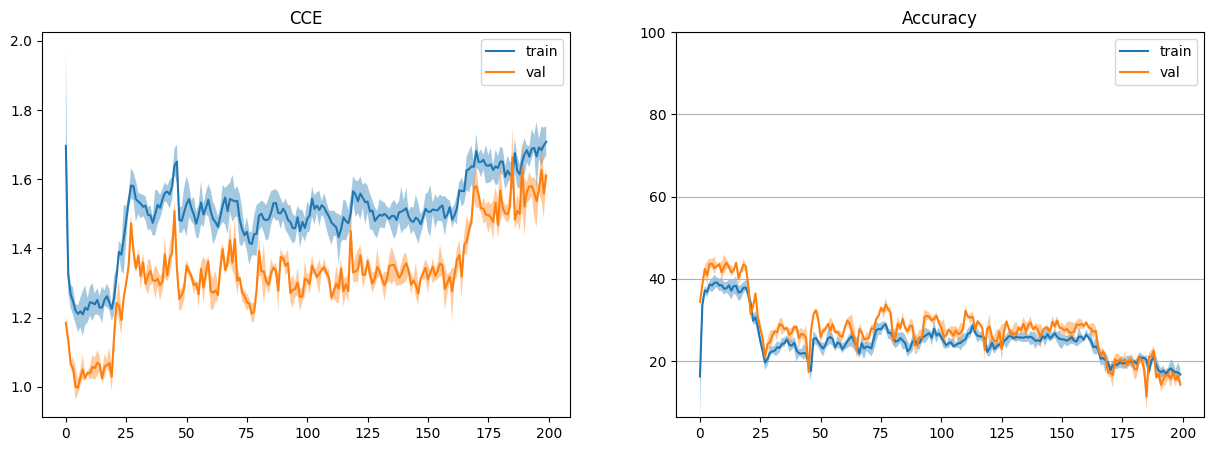


Рисунок 3 – График обучения НС с увеличенным скрытым пространством

В итоге лучший результат показала НС с базовыми параметрами.

Таблица 1 – Сравнительная таблица изменения гипперпараметров

| **Конфигурация нейросети** | **Гиперпараметры** | **Картинки** | **Комментарий** |
| --- | --- | --- | --- |
| (norm): Normalize()  (encoder): Sequential(  (0): Linear(in\_features=3072, out\_features=512, bias=True)  (1): ELU(alpha=1.0)  (2): Linear(in\_features=512, out\_features=256, bias=True)  (3): ELU(alpha=1.0)  (4): Linear(in\_features=256, out\_features=64, bias=True)  (5): Tanh()  )  (decoder): Sequential(  (0): Linear(in\_features=64, out\_features=256, bias=True)  (1): ELU(alpha=1.0)  (2): Linear(in\_features=256, out\_features=512, bias=True)  (3): ELU(alpha=1.0)  (4): Linear(in\_features=512, out\_features=3072, bias=True) | lr = 0.001,  batch\_size = 128,  epochs = 200 | C:\Users\Daniel\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\6796EAF5.tmp | Базовый вариант |
| (norm): Normalize()  (encoder): Sequential(  (0): Linear(in\_features=3072, out\_features=1024, bias=True)  (1): ELU(alpha=1.0)  (2): Linear(in\_features=1024, out\_features=512, bias=True)  (3): ELU(alpha=1.0)  (4): Linear(in\_features=512, out\_features=128, bias=True)  (5): Tanh()  )  (decoder): Sequential(  (0): Linear(in\_features=128, out\_features=512, bias=True)  (1): ELU(alpha=1.0)  (2): Linear(in\_features=512, out\_features=1024, bias=True)  (3): ELU(alpha=1.0)  (4): Linear(in\_features=1024, out\_features=3072, bias=True) | lr = 0.001,  batch\_size = 128,  epochs = 200  HIDDEN SIZE = 1024 | C:\Users\Daniel\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\E1B24A47.tmp | Изменил параметр HIDDEN SIZE = 1024 (был 512) тем самым увеличил количество нейронов |
| (norm): Normalize()  (encoder): Sequential(  (0): Linear(in\_features=3072, out\_features=512, bias=True)  (1): ELU(alpha=1.0)  (2): Linear(in\_features=512, out\_features=256, bias=True)  (3): ELU(alpha=1.0)  (4): Linear(in\_features=256, out\_features=64, bias=True)  (5): ELU(alpha=1.0)  (6): Linear(in\_features=64, out\_features=16, bias=True)  (7): Tanh()  )  (decoder): Sequential(  (0): Linear(in\_features=16, out\_features=64, bias=True)  (1): ELU(alpha=1.0)  (2): Linear(in\_features=64, out\_features=256, bias=True)  (3): ELU(alpha=1.0)  (4): Linear(in\_features=256, out\_features=512, bias=True)  (5): ELU(alpha=1.0)  (6): Linear(in\_features=512, out\_features=3072, bias=True) | lr = 0.003,  batch\_size = 128,  epochs = 200  HIDDEN SIZE = 512 | C:\Users\Daniel\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\B205F2ED.tmp | Добавил дополнительный слой в энкодер и декодер. Увеличил скорость обучения.  Стало совсем мутно и почти не разборчиво |

Визуализация проекции эмбеддинга на двумерное пространство представлена на рисунке 4.

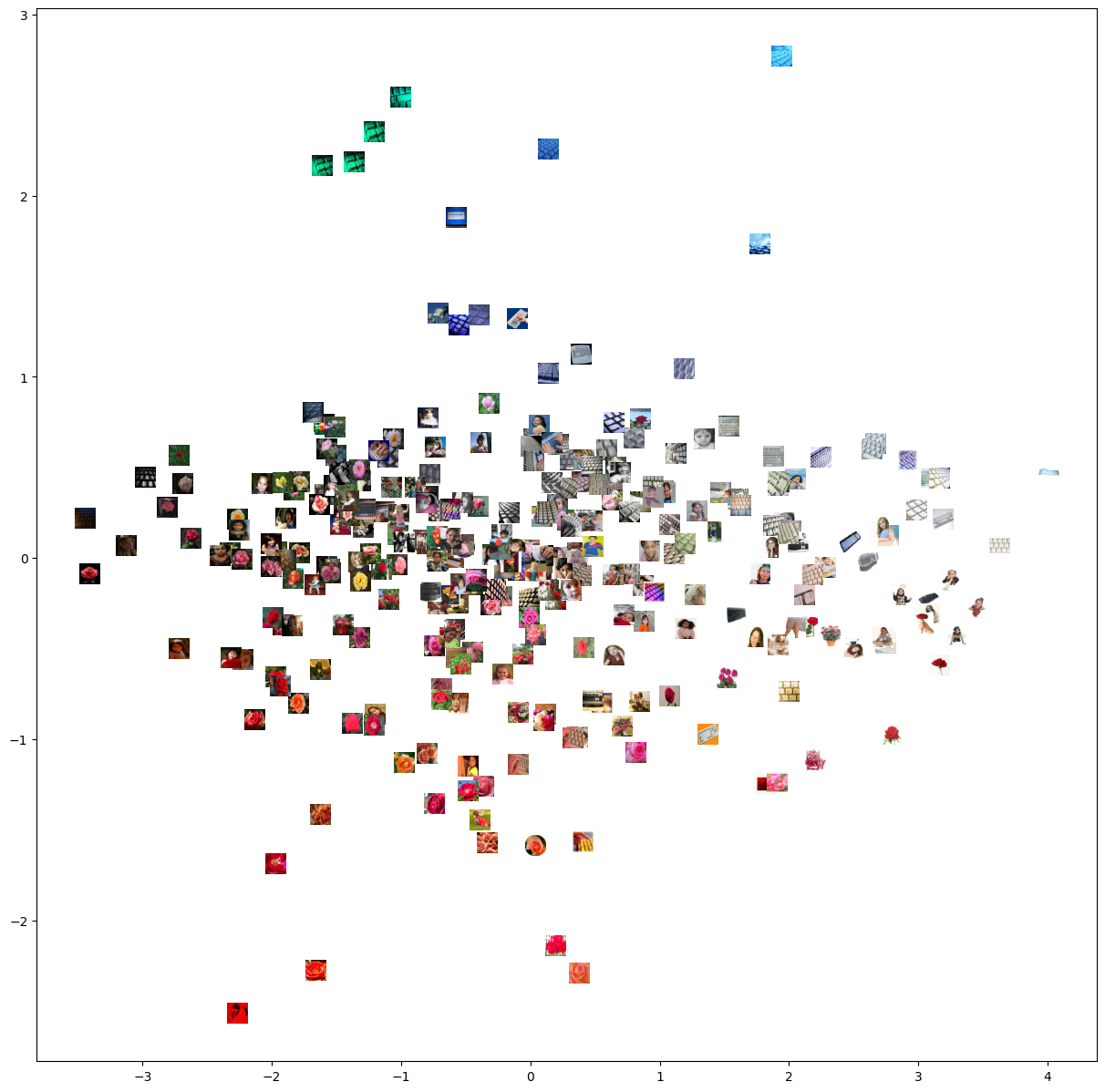


Рисунок 4 – Визуализация эмбеддинга

# Часть 2. Очистка звука

Была загруженна мелодия bogem\_rapsodia.mp3 и преобразованна в audio.wav. В качестве шума был выбран файл bee.mp3 и преобразован в noise.wav. Наложив шум по верх мелодии мы формируем нашу НС.

Здесь используется свёрточный автоэнкодер с новой функцией активации Mish, которая состоит из softmax и гипербалического тангенса.

Из НС мы получаем массив данных и с помощью прямого преобразования Фурье получаем шум.

Выход НС: (1, 2, 257) – 2 канала по 257 значений в каждом канале.

Из зашумлённой аудиозаписи мы вычитаем шум и получаем чистую аудиозапись.

В качестве сравнительного показателя здесь используется коэффициент детерминации. При чём его значение фиксируется каждую эпоху:

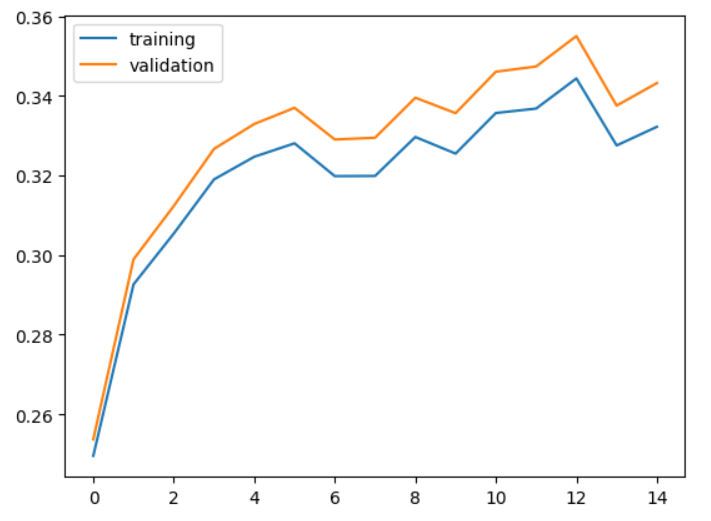
****

Рисунок 5 – График изменения коэффициента детерминации в процессе обучения

Коэффициент детерминации составляет 34,05%. И в результате мы получили аудио файл с шумом и очищенную от шума мелодию.

Прослушивание дорожки после удаления шума показало, что модель недостаточно хорошо справляется с удалением шума, т.к. шум стал менее громким, но все еще явно различим.

Далее я попробовал менять некоторые гипперпараметры, в результате получились следующие результаты:

Таблица 1 – Сравнительная таблица коэффициентов детерминации

| **Изменяемый гипперпараметр** | **Коэф. детерминации** | **График изменения коэф. детерминации** | **Комментарий** |
| --- | --- | --- | --- |
| batch\_size = 128 | 34,05% |  | Базовый вариант модели |
| batch\_size = 64 | 34% |  | Уменьшил размер батча |
| encoder:  Последний сверточный (входы = 512, выходы = 768).  decoder:  Первый сверточный (входы = 768, выходы = 512). | 33,4% |  | Уменьшил число каналов в последнем слоя энкодера с 1024 до 768, в первом слое декодера с 1024 до 768 |
| batch\_size = 256  lr = 0,000025  weight\_decay = 10-3 | 62% |  | Добавим weight\_decay = 0,001. Также используем конфигурацию НС из предыдущей модели, размер батча - 256. Скорость – 0,000025, используем SmoothL1Loss. |

Выводы по работе:

В ходе лабораторной работы был получен эмбеддинг с помощью автоэнкодера по изображениям классов из CIFAR100. Уменьшение размера батча по отношению к базовому варианту снижает качество модели, увеличение – улучшает качество модели, что совпадает с теорией. С увеличением батча снижается время обучения. Уменьшение скорости обучения и увеличение числа эпох повысило качество модели. Увеличение числа нейронов в слоях энкодера и декодера положительно влияет на качество модели, уменьшение же – наоборот, снижает качество.

Также в лабораторной работе была обучена модель автоэнкодера, нацеленная на отделение шума из аудиозаписи. Модель плохо справляется с шумовым эффектом «катер». Увеличение размера батча привело к улучшению качества модели, уменьшение – к ухудшению. Уменьшение скорости обучения и увеличение числа эпох привело к улучшению модели. Уменьшение числа каналов привело к ухудшению модели, а увеличение – к улучшению. Модель с измененной скоростью обучения и числом эпох была улучшена за счет изменения функции ошибки и добавления weight\_decay.