



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Глубокое обучение в задачах трехмерной реконструкции лиц

Рассадин Александр Георгиевич

Факультет информатики, математики и компьютерных наук

01.04.02 Прикладная математика и информатика



Содержание

- Актуальность проблемы
- Цель и задачи исследования
- Обзор работ и подходов
- Описание нейросетевого алгоритма и вычислительных экспериментов
- Результаты вычислительных экспериментов
- Оптимизация производительности нейросетевого алгоритма
- Заключение



Актуальность проблемы

Предпосылки:

- необходимость использования специализированных сенсоров имеющих высокую стоимость и ограничения на условия применимости
- отсутствие работ по восстановлению трехмерной модели лица по одному изображению методами глубокого обучения
- технология глубокого обучения наиболее эффективна в задачах компьютерного зрения, в т.ч. в задачах, связанных с идентификацией
- предлагаемая технология – пропущенное звено в современном технологическом стеке

Области применения:

- идентификация и контроль доступа
- виртуальная и дополненная реальности



Цель и задачи исследования

Целью настоящего исследования является *применение методов глубокого обучения для восстановления трехмерной модели лица по монокулярной цветной фотографии*. Необходимо также рассмотреть применимость подобной технологии на малопроизводительных платформах, как доказательство конкурентоспособности данного подхода в сравнении с использованием специализированных устройств.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- освоить технологию глубокого обучения,
- изучить подходы к предсказанию карт глубины с помощью машинного обучения,
- предложить способ решения проблемы с помощью глубоких нейронных сетей, произвести качественную и количественную оценки результата,
- изучить подходы по повышению быстродействия глубоких нейронных сетей и оптимизации потребляемых ими ресурсов,
- изучить возможность адаптации предложенного решения для работоспособности на встроенных и мобильных платформах.

Обзор работ и подходов

Восстановление карт глубины по одному цветному изображению

Liu et al – Deep Convolutional Neural Fields for Depth Estimation from a Single Image [1]

Li et al. – Depth and surface normal estimation from monocular images using regression on deep features and hierarchical [2]

Wang et al. – Towards Unified Depth and Semantic Prediction from a Single Image [3]

Mousavian et al. – Joint Semantic Segmentation and Depth Estimation with Deep Convolutional Networks [4]

Eigen et al. – Depth Map Prediction from a Single Image using a Multi-Scale Deep Network [5]

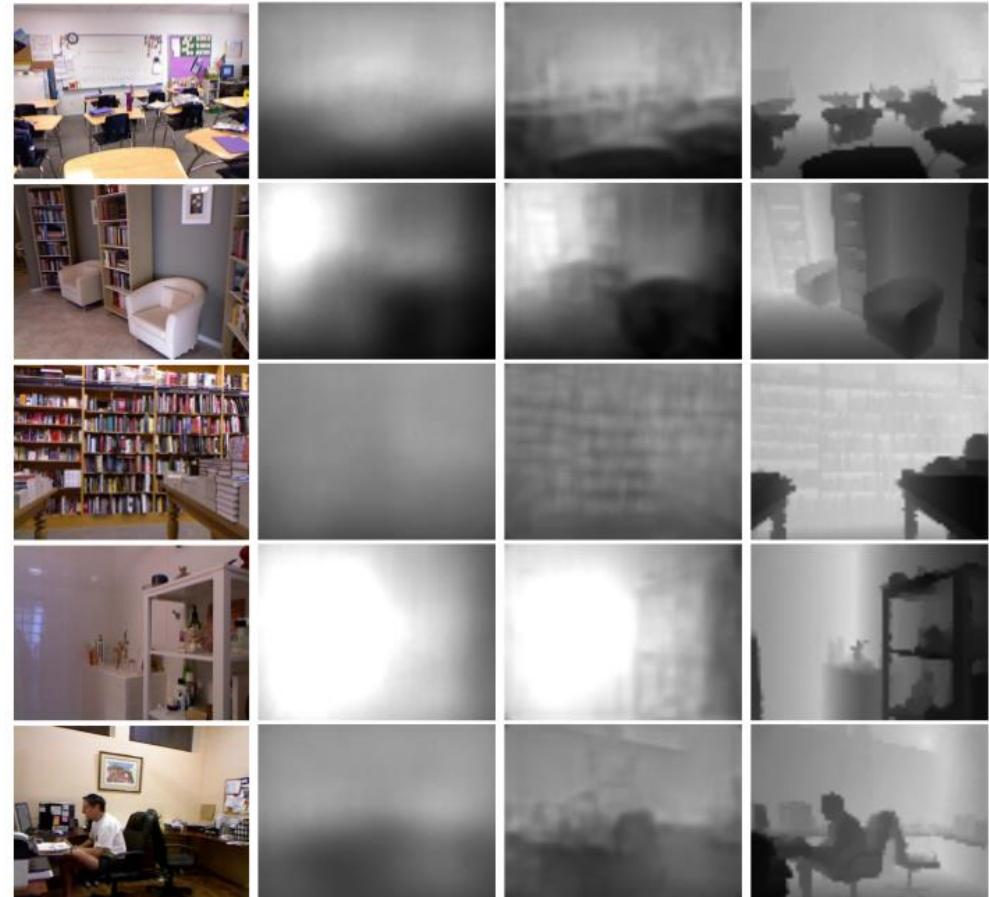


Рис. 1. Результат работы Eigen et al. [5]

Восстановление трехмерной модели лица по одному цветному изображению

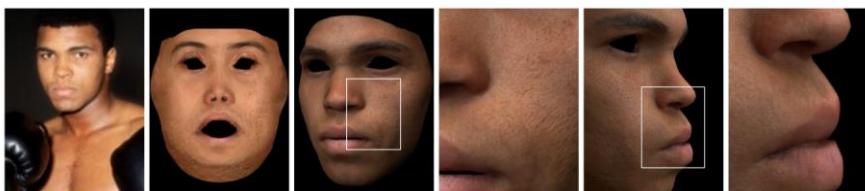


Рис. 2. Результат работы et al. [6]

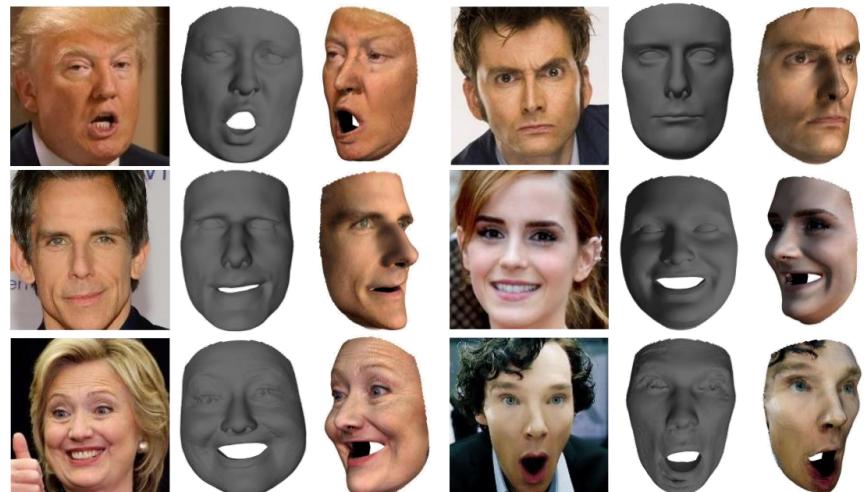
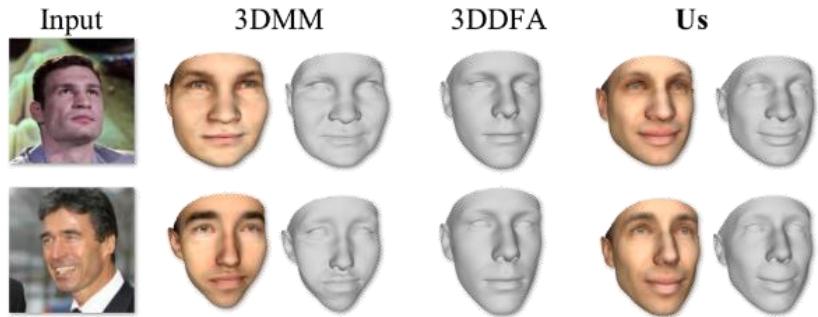


Рис. 3. Результат работы et al. [7]

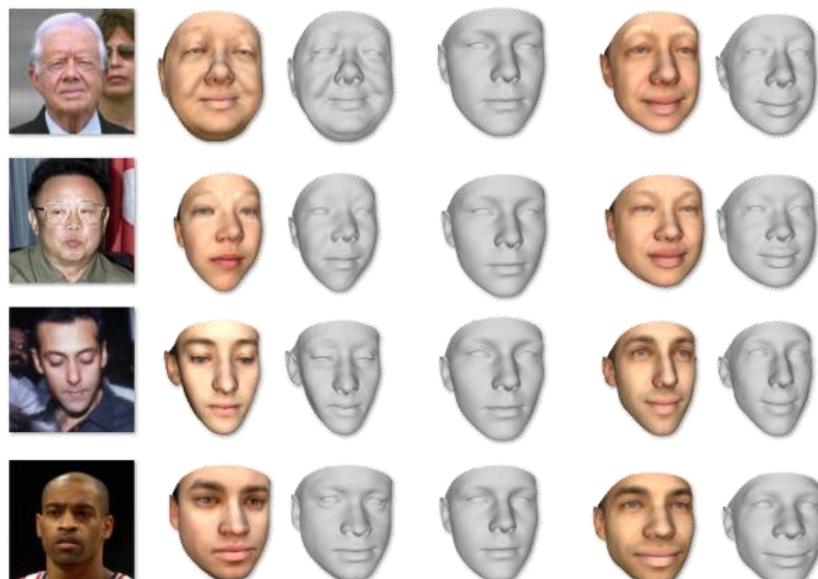


Рис. 4. Результат работы et al. [8]



Описание нейросетевого алгоритма и вычислительных экспериментов

I этап: восстановление разреженных карт глубины методом семантической сегментации.

Входом сети служат цветные изображения размера 160x160 пикселей, выход сети – разреженные карты глубины такого же размера. В качестве функции потерь использовалась категориальная кросс-энтропия. Метрика качества – доля попиксельного совпадения входа и выхода.

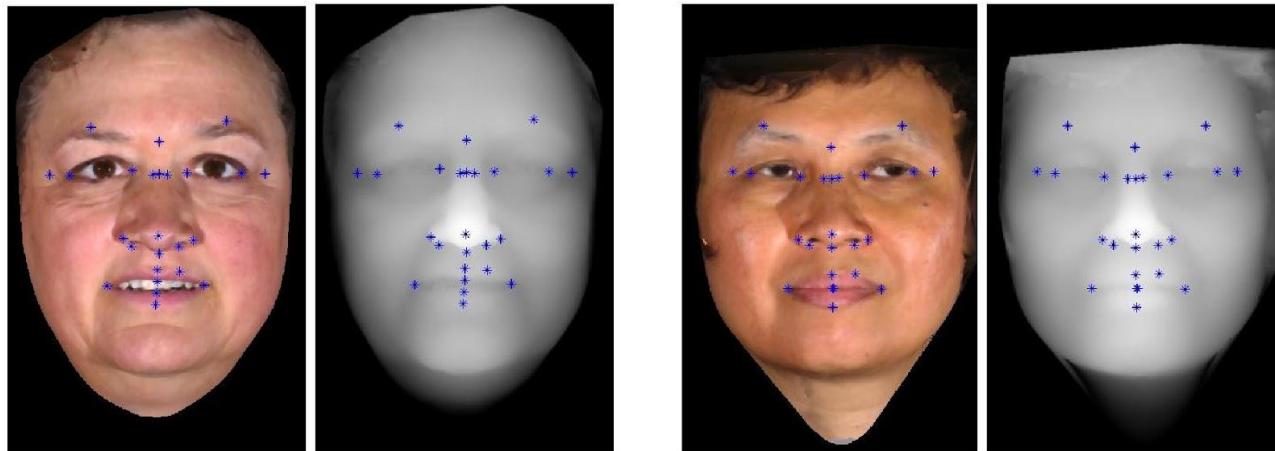
II этап: регрессия оригинальных карт глубины.

Входом сети служат двухканальные изображения размера 160x160 пикселей, где один канал – оригинальные черно-белые изображения, другой – предсказания сети с предыдущего этапа. Выход сети – карты глубины субмиллиметровой точности размером 80x80. Метрика качества – средняя абсолютная разница между входом и выходом.

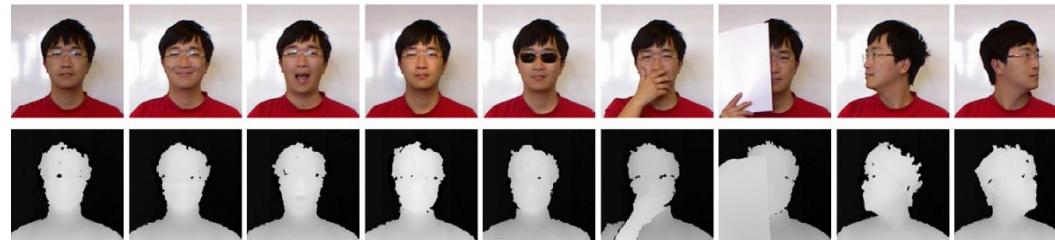
3D Basel Face Model [9]



Texas 3D Face Recognition Database [10] – [12]

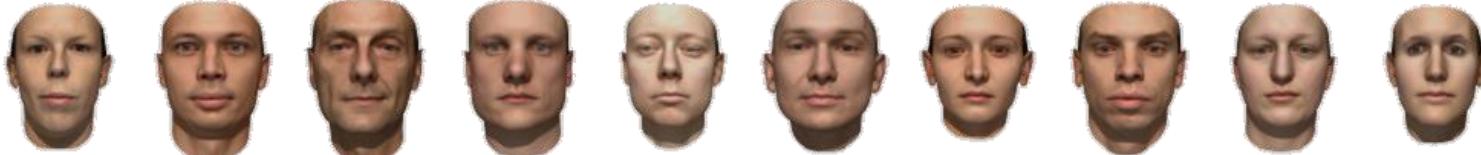


EURECOM Kinect Face Dataset [13]

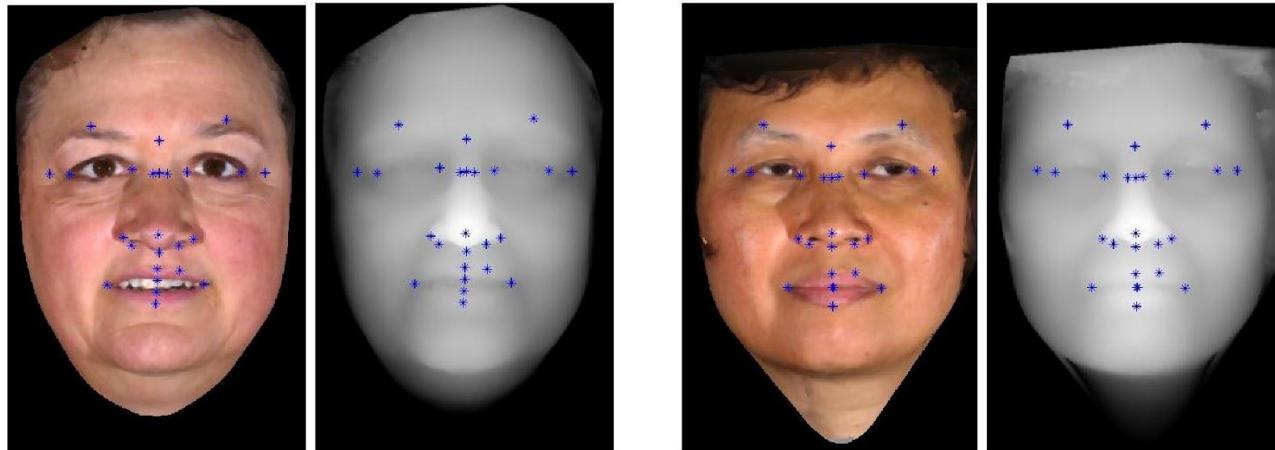


Biwi Kinect Head Pose Database [14], Florence Superface dataset [15]

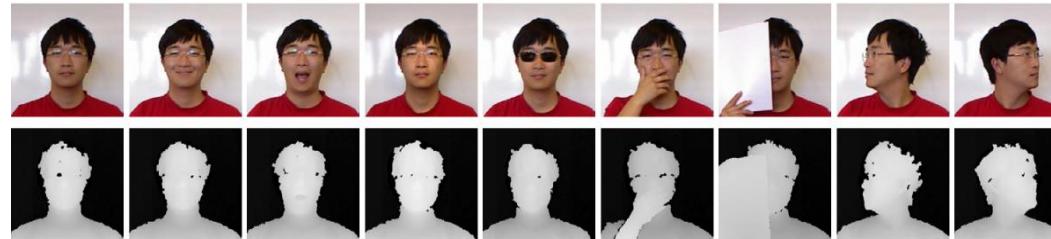
3D Basel Face Model [9]



Texas 3D Face Recognition Database [10] – [12]



EURECOM Kinect Face Dataset [13]



Biwi Kinect Head Pose Database [14], Florence Superface dataset [15]

Результаты вычислительных экспериментов

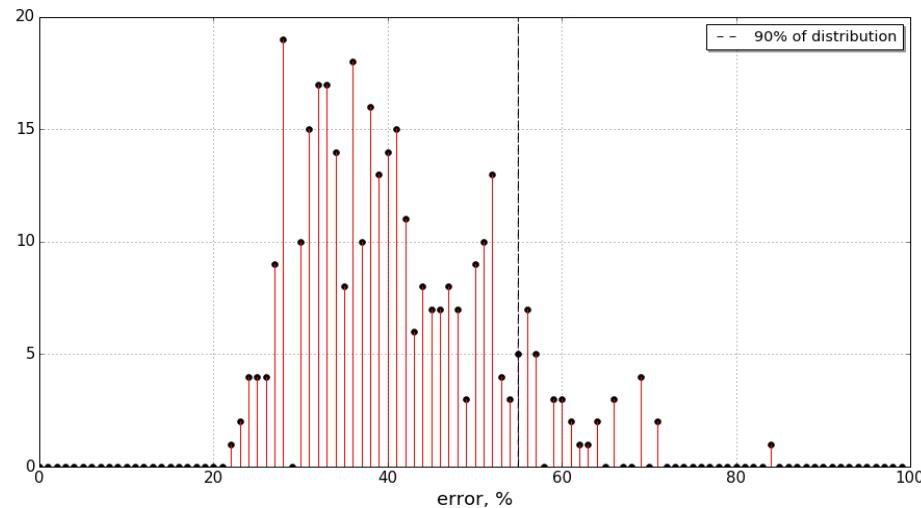


Рис. 5. Гистограмма ошибки для top1-классов на тестовой выборке

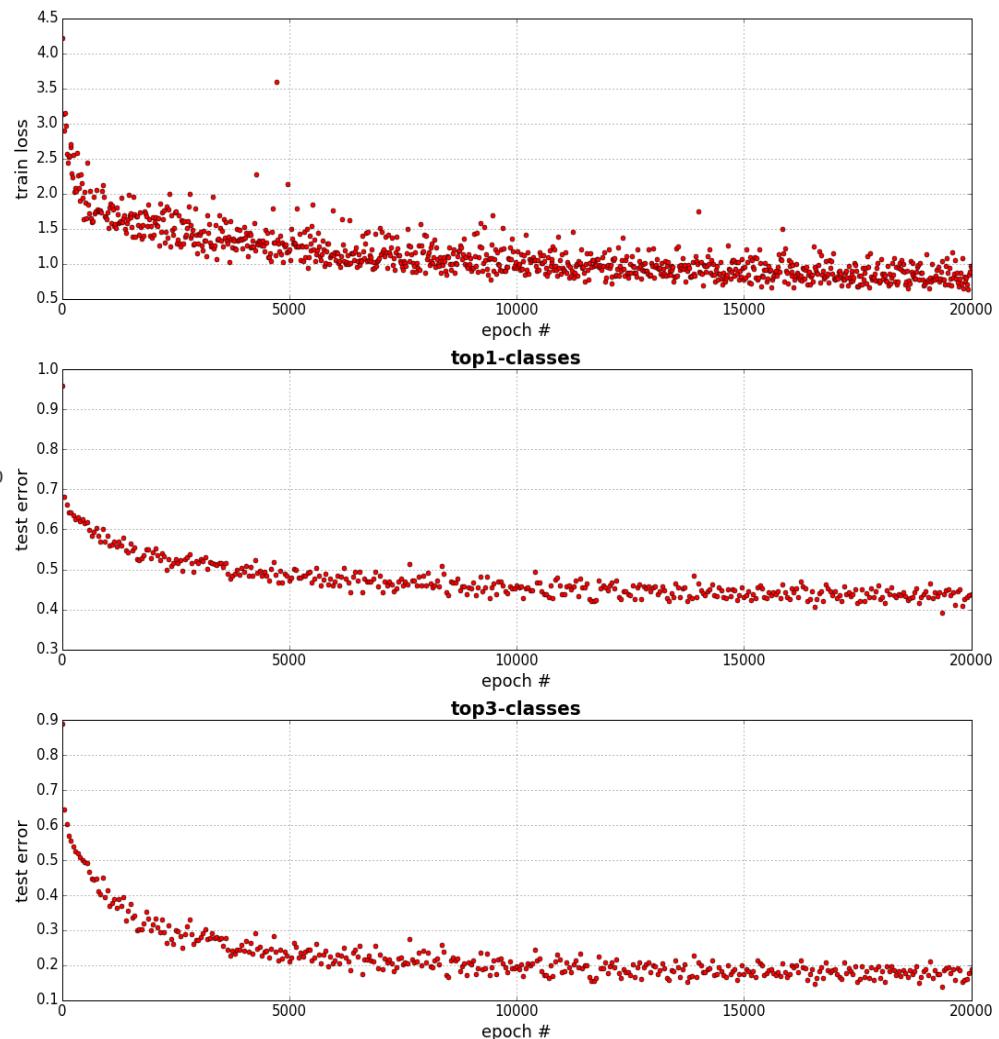


Рис. 6. Изменение характеристик вычислительного эксперимента в зависимости от итерации обучения: верхний ряд – функция потерь на обучающей выборке, средний ряд – точность для top1-классов тестовой выборки, нижний ряд – точность для top3-классов тестовой выборки

величина ошибки, %

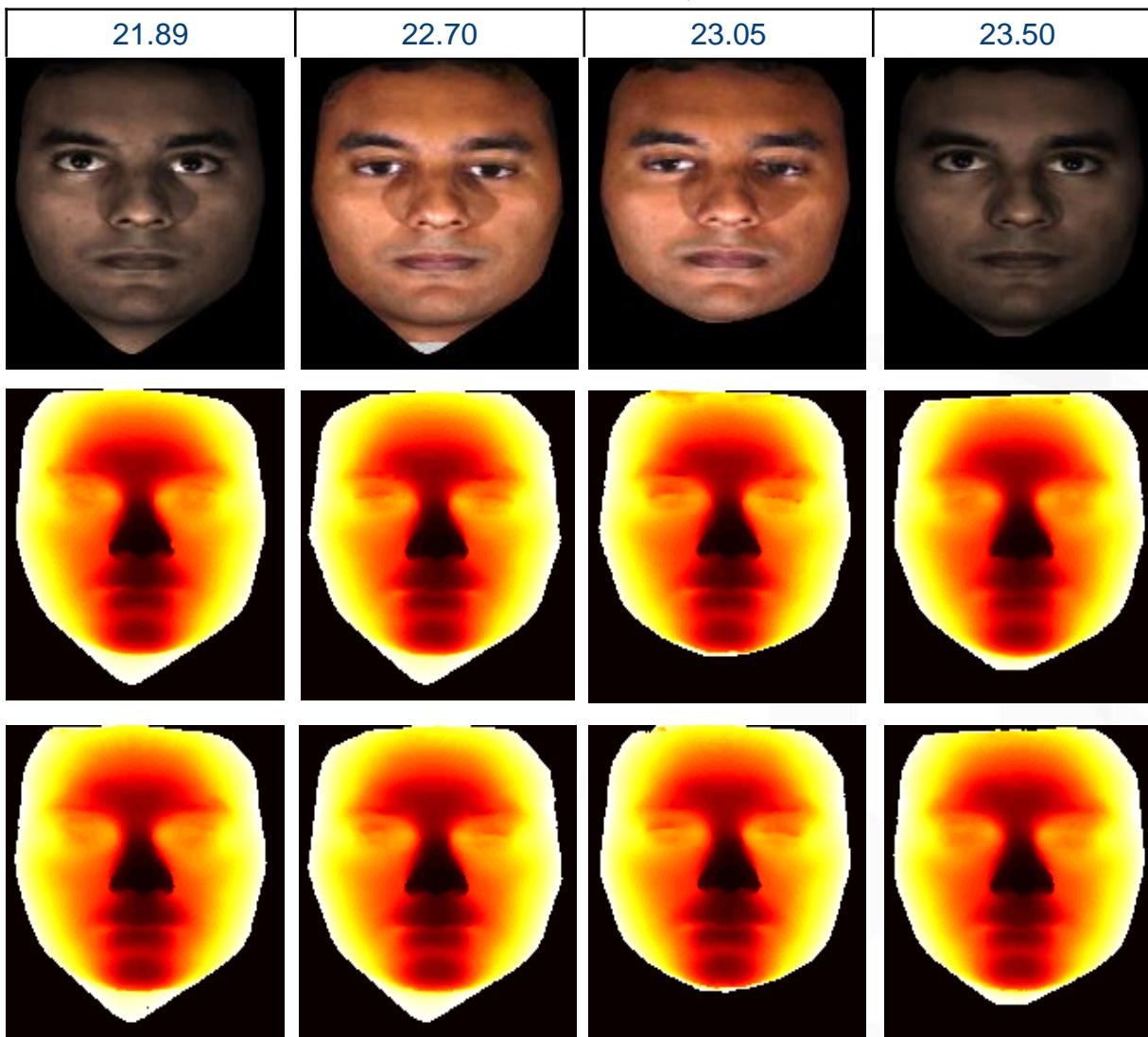


Рис. 7. Четыре лучших отклика сети на тестовой выборке: верхний ряд – оригинальное цветное изображение, средний ряд – оригинальная карта глубины, нижний ряд – результат предсказания модели

величина ошибки, %

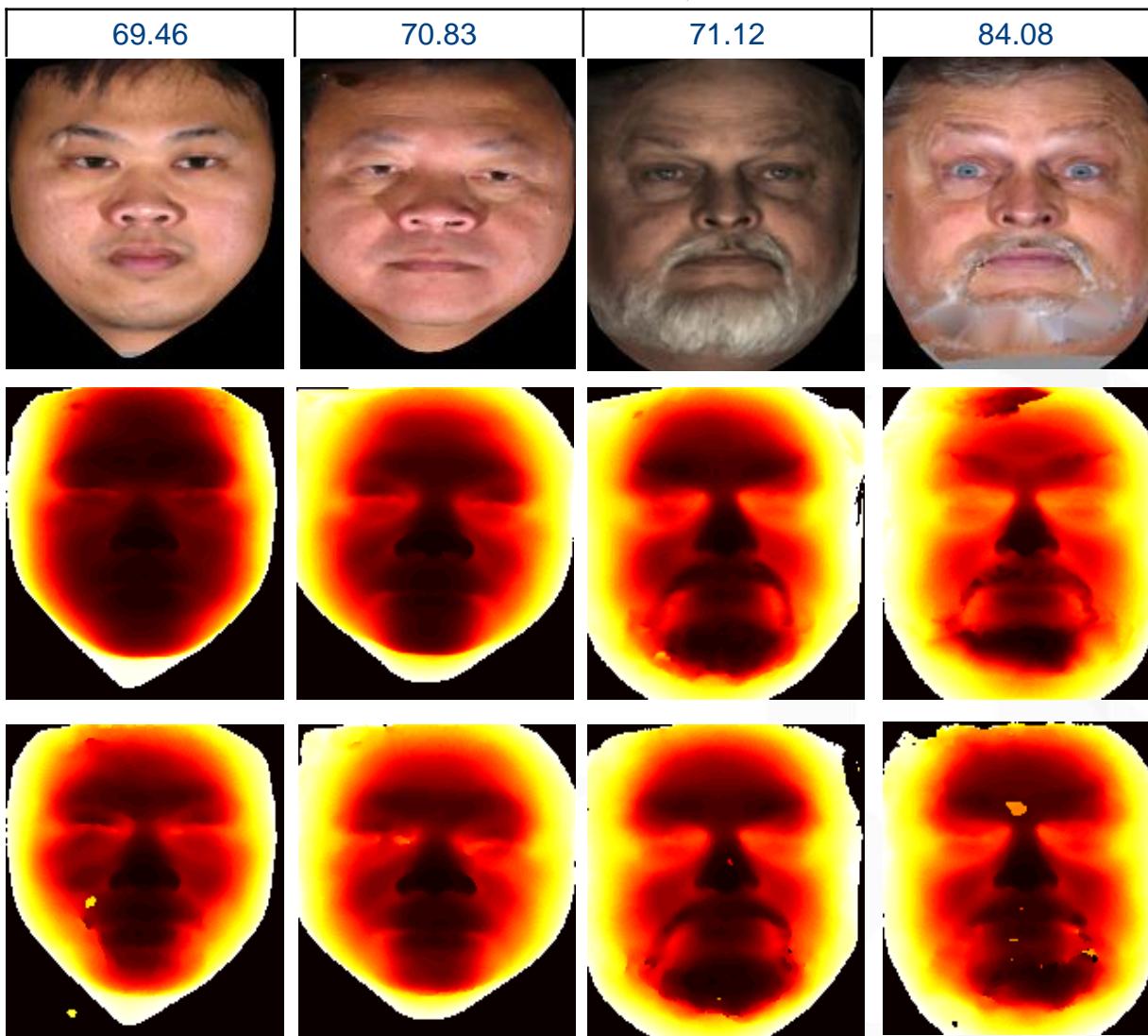


Рис. 8. Четыре худших отклика сети на тестовой выборке: верхний ряд – оригинальное цветное изображение, средний ряд – оригинальная карта глубины, нижний ряд – результат предсказания модели

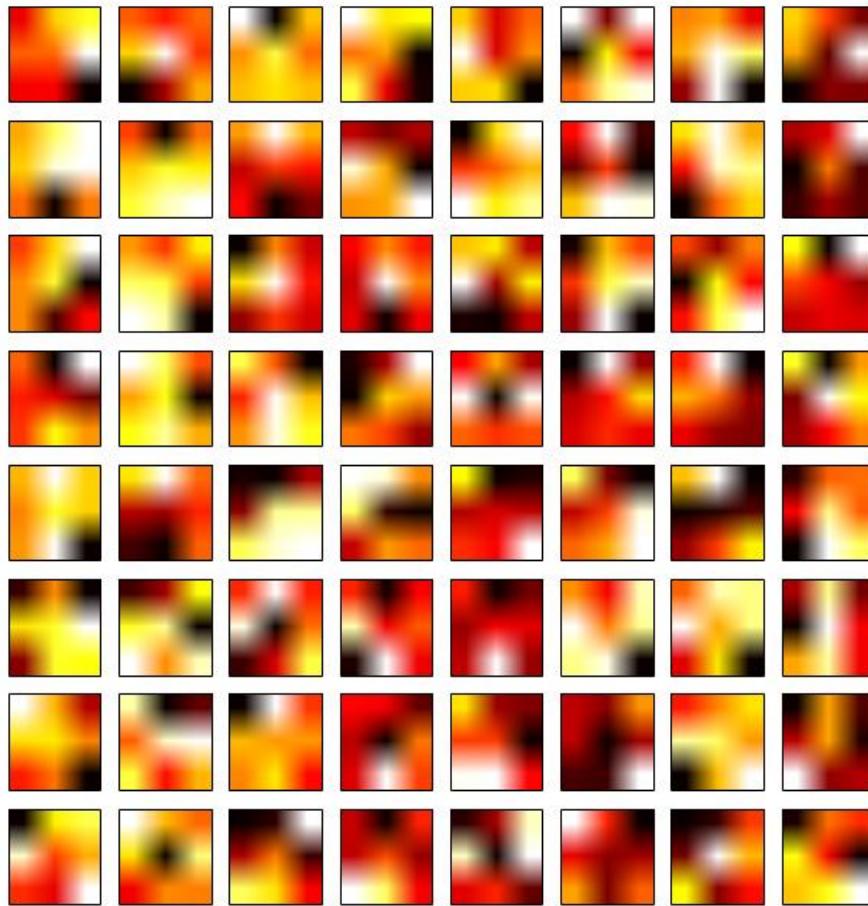


Рис. 9. Карты фильтров последнего слоя сети

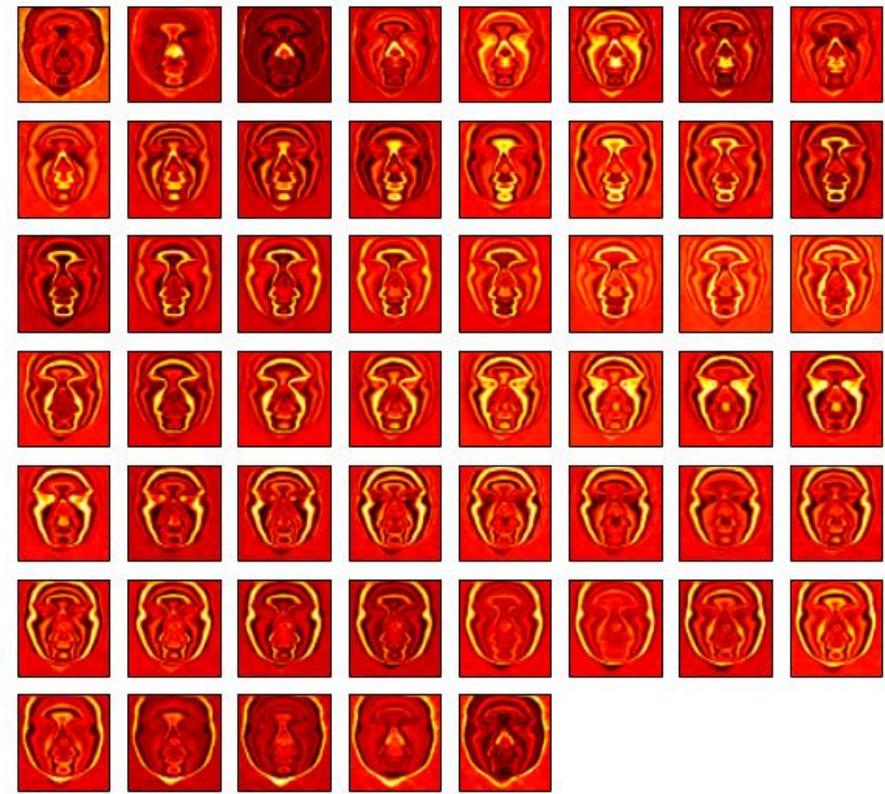


Рис. 10. Карты откликов последнего слоя сети

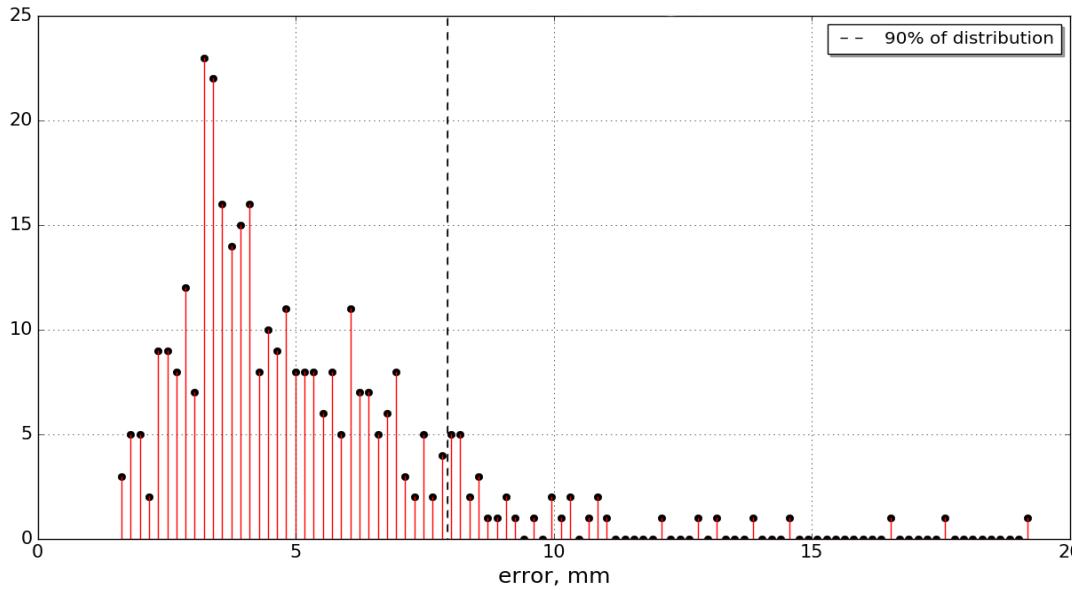


Рис. 11. Гистограмма ошибки на тестовой выборке

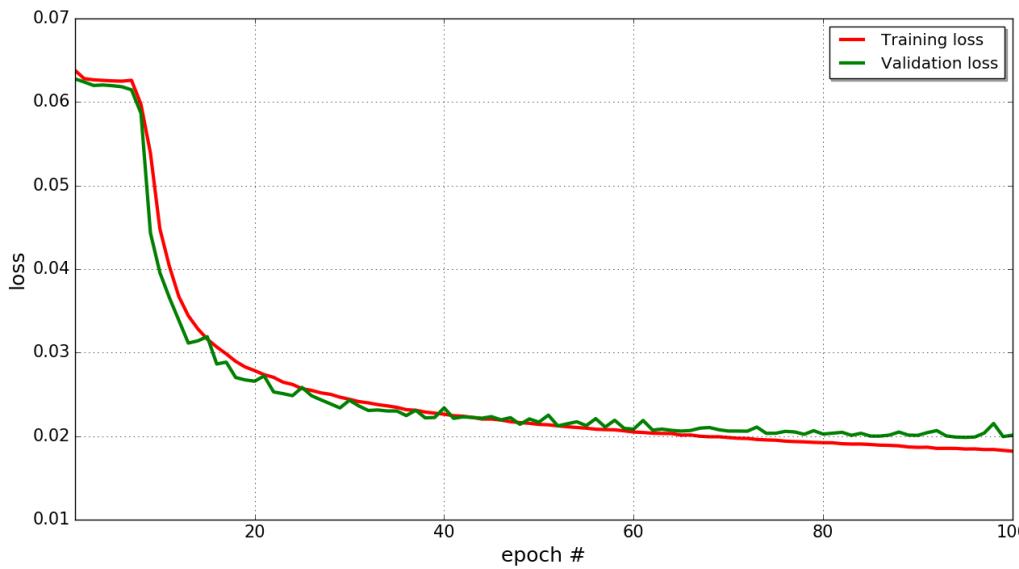


Рис. 12. Изменение характеристик вычислительного эксперимента в зависимости от итерации обучения

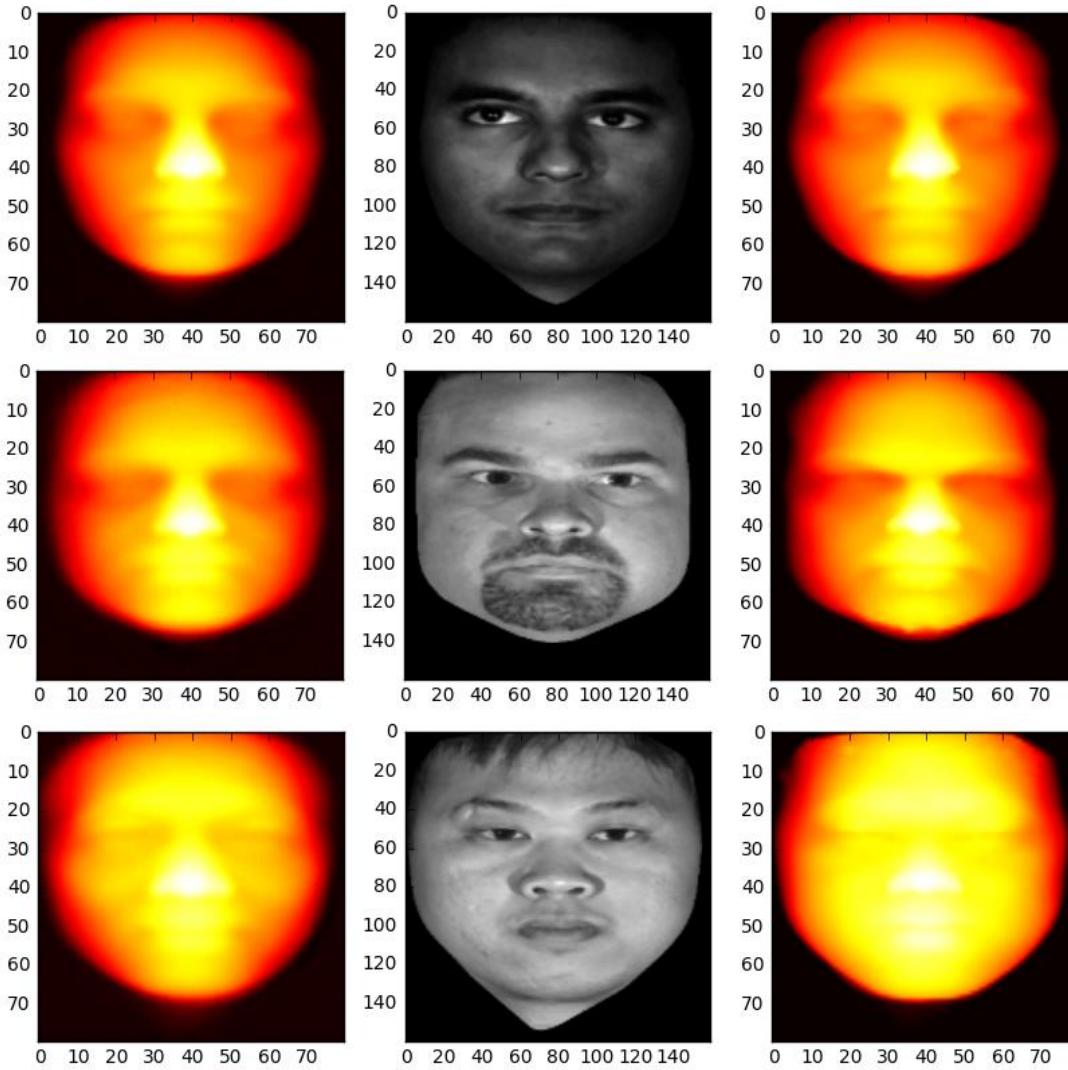


Рис. 13. Три результата предсказания модели: верхний ряд – лучшее предсказание, средний ряд – предсказание, соответствующее среднему уровню ошибки, нижний ряд – худшее предсказание; слева – предсказание модели, справа – фактическая карта глубины

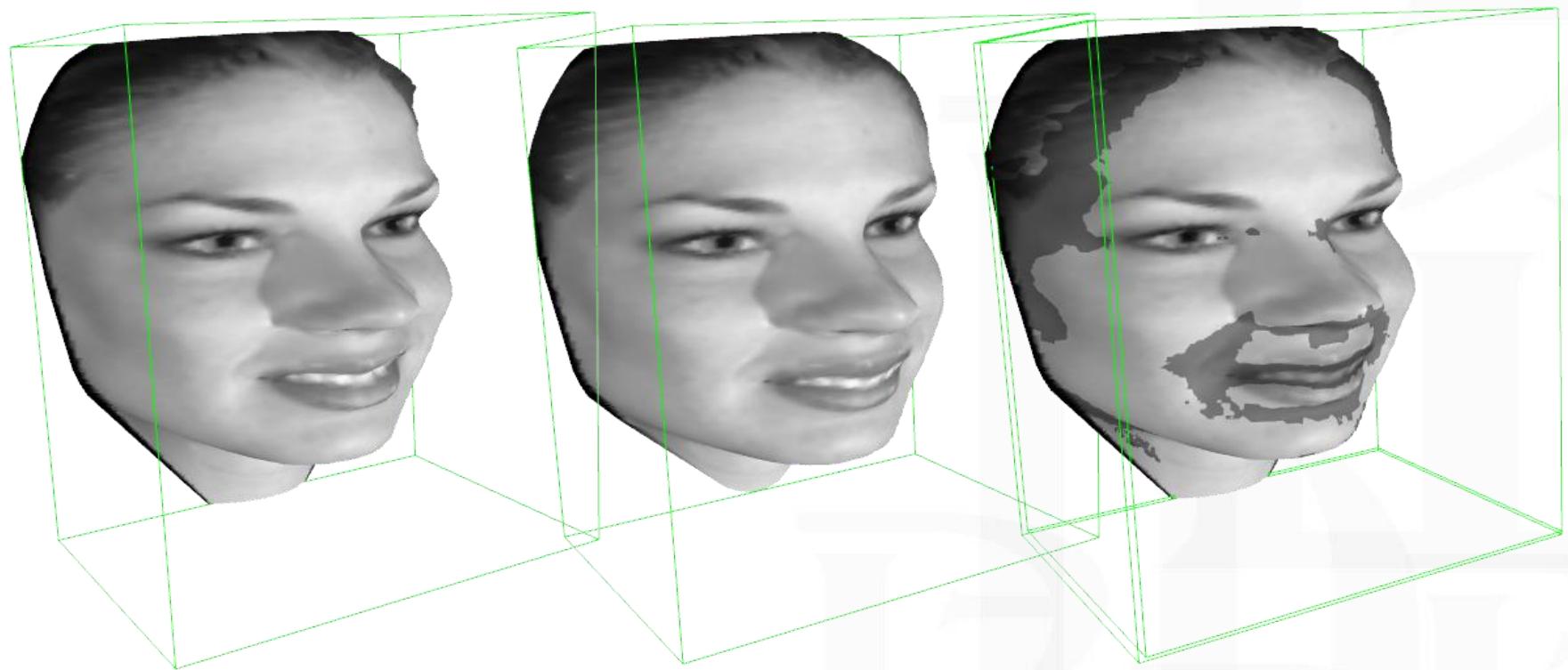


Рис. 14. Слева направо: предсказание нейросетевой модели, оригинальная трехмерная модель, визуализация ошибки (тёмные области)



Оптимизация производительности нейросетевого алгоритма

В ходе решения задачи, был сделан обзор литературы методов оптимизации производительности нейронных сетей и были найдены следующие методы:

- Прунинг
- Передача знаний (distillation the knowledge)
- Хеширование и квантизация
- Разложение матриц (тензорное разложение)
- Бинаризация весов
- Архитектурная оптимизация

	Память (обучение)	Память (исполнение)	Ускорение исполнения
Deep Compression [16]	нет	49	неизвестно
FitNets [17]	нет	36	13.36
HashedNets [18]	нет	64	неизвестно
CP-Decomposition [19]	нет	12	4.5
TensorNet [20]	неизвестно	80	неизвестно
BinaryNet [21]	~32 (теоретическое)	~32 (теоретическое)	от 3.4 до 23
Binary-Weight-Network and XNOR-Net [22]	нет	67	58 (CPU)
SqueezeNet [23]	неизвестно	50	1.
Tiny Darknet [24]	неизвестно	60	2.9
BranchyNet [25]	нет	нет	1.9

Таблица 1. Численные характеристики методов оптимизации.

	Память (обучение)	Память (исполнение)	Ускорение исполнения
Deep Compression [51]	нет	49	неизвестно
FitNets [54]	нет	36	13.36
HashedNets [54]	нет	64	неизвестно
CP-Decomposition [56]	нет	12	4.5
TensorNet [59]	неизвестно	80	неизвестно
BinaryNet [60]	~32 (теоретическое)	~32 (теоретическое)	от 3.4 до 23
Binary-Weight-Network and XNOR-Net [61]	нет	67	58 (CPU)
SqueezeNet [63]	неизвестно	50	1.
Tiny Darknet [64]	неизвестно	60	2.9
BranchyNet [65]	нет	нет	1.9

Таблица 1. Численные характеристики методов оптимизации.

	Время обучения, мс	Время выполнения, мс	Размер модели, Мб	Точность, %
VGG-S (baseline)	43.7	33.4	372.2	97.13
SqueezeNet-1.1 (baseline)	22.94	4.94	2.8	89.14
CP-Decomposition	22.94	7.74	2.1	87.5
HashedNets	294.8	158.2	68.3	96.31
Binary-Weight-Network	83.8	33.5	11.6	98.57
XNOR-Net	84.3	34.2	11.6	58.81
XNOR-Net w/o weights activation	43.4	34.1	11.6	88.32

Таблица 2. Результаты численных экспериментов.

Выводы.

- Методы на основе бинаризации весов показали практическую несостоительность. Текущее развитие программно-аппаратного комплекса не позволяет полностью раскрыть потенциал этого подхода. Применение таких технологий на мобильных платформах в ближайшее время является, по крайней мере, трудоемким процессом.
- Методы на основе тензорного разложения просты в имплементации и показывают стабильный выигрыш в производительности при незначительной потере точности. Основное преимущество их в том, что они не требуют реализации на целевой (мобильной) платформе.
- Методы на основе квантизации и хеширования гарантируют высокую оптимизацию потребления памяти. Реализация их на мобильной платформе намного менее трудоемкая, чем в случае бинаризационных методов. Скорость выполнения, однако, может сильно меняться в зависимости от платформы.



Заключение

Поставленная цель – применение методов глубокого обучения для восстановления трехмерной модели лица по одной монокулярной цветной фотографии выполнена в полном объеме. Результатом ее выполнения является уникальный и эффективный численный алгоритм. Дополнительно рассмотрен вопрос о применимости данного алгоритма на малопроизводительных платформах, что потенциально расширяет области его применения до таких устройств, как мобильные телефоны или портативные камеры. Массовое распространение такой технологии может произвести настоящий прорыв в таких областях, как видеонаблюдение (идентификация и контроль доступа) и виртуальная реальность.

Список использованной литературы.

1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8
9. 9
10. 10
11. 11
12. 12
13. 13
14. 14
15. 15
16. 16
17. 17
18. 18
19. 19
20. 20
21. 21
22. 22
23. 23
24. 24
25. 25





НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Спасибо
за внимание!