



СРАВНИТЕЛЬНЫЙ АНАЛИЗ И ВЫБОР АРХИТЕКТУРЫ ГЛУБОКОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ГЕНЕРАЦИИ ЭМБЕДДИНГОВ В АЛГОРИТМАХ РАСПОЗНАВАНИЯ ЛИЦ ПРИ УЧЕТЕ ПОСЕЩАЕМОСТИ

Абдуллаев Дилавар Бахтиярович

студент группы 304-25

направления «Компьютерные сети и их программное обеспечение»

Ташкентского университета

информационных технологий имени Мухаммада аль-Хорезми

Аннотация. Данная статья посвящена сравнительному анализу ключевых архитектур глубоких нейронных сетей, предназначенных для генерации эмбеддингов (векторных представлений) лиц, с акцентом на их применимость в системах автоматизированного учета посещаемости. В условиях растущего спроса на высокоточные, надежные и быстрые системы биометрической идентификации, актуальность выбора оптимальной архитектуры для данной задачи является критически высокой. В работе исследуются и сопоставляются две доминирующие на данный момент архитектуры: FaceNet (использующая Triple Loss) и ArcFace (основанная на Arc Margin Loss), а также упоминаются другие перспективные подходы. Анализ фокусируется на точности, устойчивости к вариациям (поза, освещение) и вычислительной эффективности, необходимой для работы в режиме реального времени в системах контроля доступа и учета.

Ключевые слова: Распознавание лиц (Face Recognition), Глубокие нейронные сети (Deep Neural Networks), Эмбеддинги (Embeddings), FaceNet, ArcFace, Учет посещаемости.

ВСТУПЛЕНИЕ

Распознавание лиц стало одной из наиболее успешных и широко применяемых областей компьютерного зрения, находящей свое применение от систем безопасности до мобильной аутентификации. Особое место занимает использование этой технологии для автоматизированного учета посещаемости сотрудников, студентов или посетителей. Ключевым этапом в любом современном алгоритме распознавания лиц является генерация эмбедингов — компактных, высокодискриминационных векторных представлений, которые позволяют количественно оценить сходство между разными изображениями одного и того же лица.

Эффективность всей системы напрямую зависит от качества этих эмбедингов. Исторически сложилось несколько ключевых архитектур глубоких нейронных сетей, решающих эту задачу. Среди них особо выделяются FaceNet и ArcFace.

Цель данной статьи — провести детальный сравнительный анализ архитектур FaceNet и ArcFace (и других, если применимо), определить их фундаментальные различия, оценить их производительность в контексте требований к системе учета посещаемости (высокая точность, скорость идентификации) и дать обоснованные рекомендации по выбору оптимальной архитектуры.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ АРХИТЕКТУР ДЛЯ ГЕНЕРАЦИИ ЭМБЕДИНГОВ

Архитектура FaceNet, представленная Google в 2015 году, стала прорывом в области распознавания лиц. Она использует глубокую сверточную нейронную сеть (часто Inception или ResNet) для прямого отображения изображения лица в низкоразмерное Евклидово пространство (эмбединг, обычно 128 или 512 измерений).

Ключевая особенность FaceNet — это функция потерь Triple Loss (Тройная функция потерь). Она нацелена на минимизацию расстояния между эмбедингами изображений одного и того же человека (Anchor и Positive) и



максимизацию расстояния до эмбедингов изображений другого человека (Negative). Формально:

$$||E(A) - E(P)||^2 + \alpha < ||E(A) - E(N)||^2$$

Где — $E(\cdot)$ эмбединг, а α — маржа, гарантирующая, что эмбединги разных людей будут находиться на достаточном расстоянии друг от друга.

- Плюсы для учета посещаемости: Высокая точность, компактные эмбединги.

- Минусы для учета посещаемости: Сложность выбора "сложных троек" (hard triplets) во время обучения, что может замедлять сходимость и усложнять тренировку.

ArcFace (представленная в 2019 году) является более современным и на данный момент, в большинстве случаев, более эффективным решением. Вместо того чтобы работать в Евклидовом пространстве, ArcFace фокусируется на угловом (косинусном) пространстве, используя Additive Angular Margin Loss (AAML), или просто Arc Margin Loss.

ArcFace преобразует традиционную классификационную функцию потерь (Softmax Loss) в угловую маржинальную функцию потерь, которая увеличивает угловое расстояние между эмбедингом лица и весами соответствующего ему класса. Это приводит к тому, что лица одного человека образуют более плотный кластер на гиперсфере, а кластеры разных людей становятся более различимыми (дискриминационными).

- Плюсы для учета посещаемости: Превосходная дискриминационная способность, более четкое разделение классов, лучшая устойчивость к вариациям освещения и позы. Упрощенное обучение по сравнению с FaceNet.

- Минусы для учета посещаемости: Требуется тщательной настройки гиперпараметров.

В контексте учета посещаемости, система должна обладать двумя критическими качествами: минимальным количеством ложных допусков (False Accept Rate, FAR) и высокой скоростью распознавания.

Таблица 1

Сравнительный анализ и выбор архитектуры

| Характеристика | FaceNet | ArcFace | Примечание |
|-------------------------------|---------------------------------------|---|---|
| Основной принцип | Triple Loss (Евклидово расстояние) | Arc Margin Loss (Угловое расстояние) | Угловое разделение ArcFace лучше для четкого разделения "свой/чужой". |
| Точность (LFW/MegaFace) | Высокая | Выше | ArcFace демонстрирует лучшие результаты на сложных наборах данных. |
| Дискриминационная способность | Хорошая | Отличная | Эмбединги ArcFace более плотные и лучше различимы. |
| Сложность обучения | Высокая (выбор "троек") | Ниже (стандартная классификационная задача) | ArcFace быстрее сходится и легче в реализации. |



| Вычислительная эффективность | Зависи т от базовой CNN (например, Inception) | Зависит от базовой CNN (например, ResNet, MobileNet) | Обе модели могут быть оптимизирова ны для скорости с помощью легковесных бэкбонов. |
|---------------------------------|---|--|--|
|---------------------------------|---|--|--|

ВЫВОД

Для систем учета посещаемости, где критически важна высокая точность идентификации и минимизация ложных срабатываний при повседневных вариациях (плохое освещение, небольшие повороты головы), архитектура ArcFace с соответствующим угловым маргинальным лоссом демонстрирует превосходную эффективность. Она обеспечивает более четкое разделение лиц разных людей, что напрямую влияет на надежность системы. Современные системы распознавания лиц для учета посещаемости требуют не просто высокой точности, а бескомпромиссной надежности в реальных условиях эксплуатации. Анализ показал, что эволюция от FaceNet к ArcFace ознаменовала переход к более эффективному методу обучения, ориентированному на угловое разделение классов на гиперсфере.

ArcFace, благодаря своему Additive Angular Margin Loss, является наиболее предпочтительной архитектурой для генерации эмбеддингов в системах учета посещаемости. Она гарантирует, что лица одного человека будут иметь высокую схожесть (малое угловое расстояние), а лица разных людей будут иметь максимально низкую схожесть, что является ключом к низкому показателю ложных допусков (FAR). Дальнейшие исследования должны быть сосредоточены на оптимизации ArcFace-подобных моделей с использованием легковесных базовых сетей (например, MobileNetV2/V3) для

обеспечения высокой скорости работы на встроенных или мобильных устройствах, используемых в турникетах и системах доступа.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. Schroff, F., Kalenichenko, D., & Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
2. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
3. Beibei Xu, Wensheng Wang, Leifeng Guo, Guipeng Chen, Yongfeng Li, Zhen Cao, Saisai Wu, CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss, *Computers and Electronics in Agriculture*, Volume 193, 2022, 106675, ISSN 0168-1699.
4. Joseph A. Mensah, Justice K. Appati, Elijah K.A Boateng, Eric Ocran, Louis Asiedu, FaceNet recognition algorithm subject to multiple constraints: Assessment of the performance, *Scientific African*, Volume 23, 2024, e02007, ISSN 2468-2276.
5. Lingling Yang, Xingshi Xu, Jizheng Zhao, Huaibo Song, Fusion of RetinaFace and improved FaceNet for individual cow identification in natural scenes, *Information Processing in Agriculture*, Volume 11, Issue 4, 2024, Pages 512-523, ISSN 2214-3173.