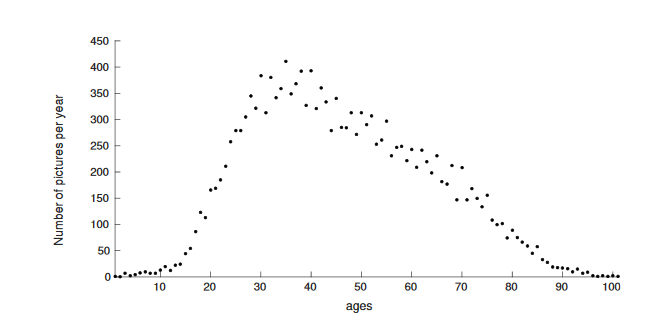
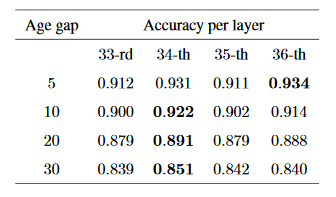
1. [AgeDB](http://openaccess.thecvf.com/content_cvpr_2017_workshops/w33/papers/Moschoglou_AgeDB_The_First_CVPR_2017_paper.pdf)

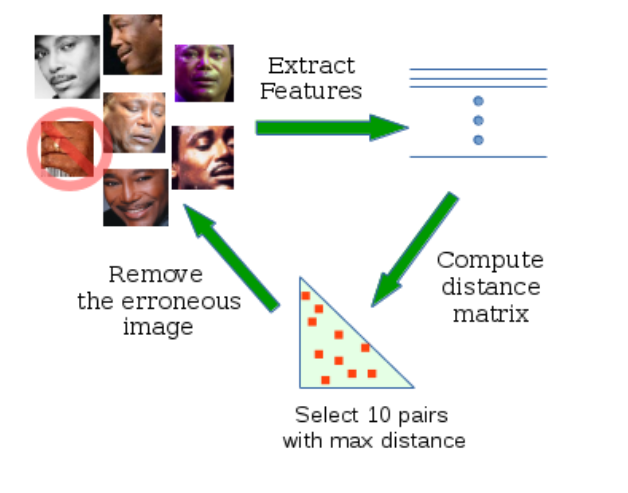
Набор изображений лиц людей разных возрастов в естественных условиях. Скачать его можно [здесь](https://ibug.doc.ic.ac.uk/resources/agedb/). 

Этот набор может быть использован в инвариантных возрасту экспериментах по распознаванию лиц в реальных условиях: чувствительность алгоритма распознавания лица может быть измерена при увеличении возрастного разрыва между изображениями одного и того же человека. Возрастные метки точно соответствуют, так как AgeDB собран *вручную*. Более того, все люди были зафиксированы бесконтрольно, в реальных условиях, то есть имеют разные позы, эмоции на лице, могут быть даже частично загорожены. С целью проведения подобных экспериментов на основе AgeDB была создана база данных LFW: изображения разбиты согласно предопределённому значению возрастной разницы (5, 10, 20 и 30 лет)

1 Инвариантное возрасту распознавание лиц с использованием глубокой сети VGG Face

1. [UmdFaces](https://arxiv.org/pdf/1611.01484.pdf)

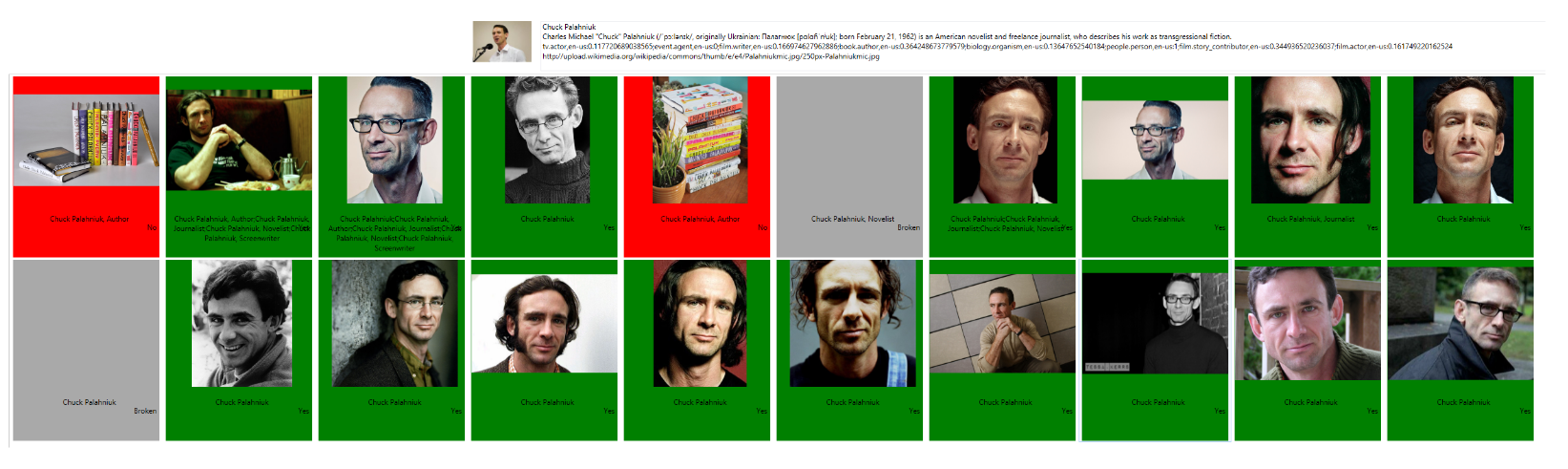
Этот набор был пропущен через свёрточную нейронную сеть и затем снабжён разного рода информацией: рамки, выражение лица, гендерная классификация, которые затем были проверены людьми и скорректированы. Из 115000 изображений, проанализированных двумя экспертами, 28084 содержали неверную информацию. Тем не менее этого анализа оказалось недостаточно, и набор подвергся дополнительной проверке.

2 Стратегия окончательной очистки набора данных.

В результате такой очистки были удалены ещё 12789 изображений. В результате он состоит из 367888 изображений для 8277 людей, разделённых на непересекающиеся тренировочные и тестовые части.

1. [MS-Celeb-1M](https://arxiv.org/pdf/1607.08221.pdf)

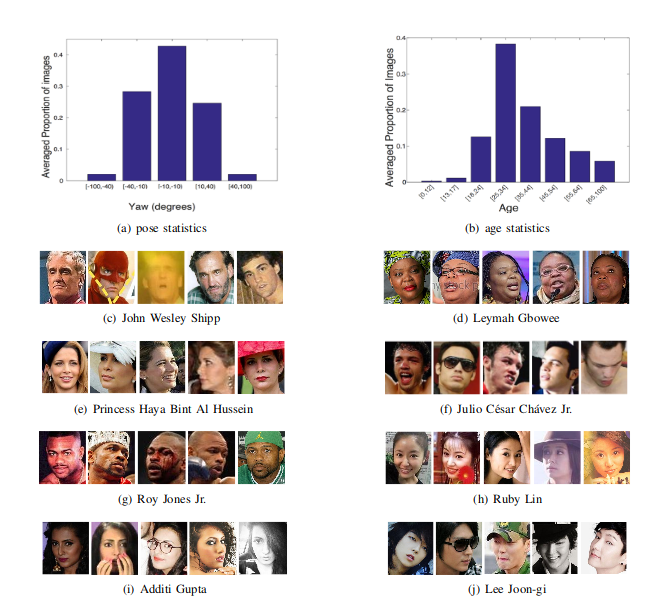
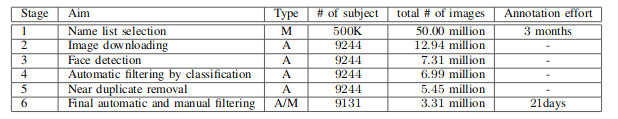
Набор, состоящий из 10 миллионов изображений лиц знаменитостей. Он используется больше для поиска похожих изображений, а не для точности распознавания лица. Другими словами, для оценки метрики расстояния между похожими лицами. Он намеренно включает в себя лица похожих друг на друга знаменитостей (и даже близнецов), а также очень разные изображения одного человека (разный возраст, наличие макияжа, проведённые пластические операции, в том числе по смене пола).

Выходные данные, получаемые при использовании этого датасета, более близки к привычной человеческой оценке, так как распознанное лицо затем связывается с соответствующей о нём информацией в базе знаний (если распознали известную актрису, можно будет узнать не только её имя, но и также сам факт того, что она является актрисой, фильмы, где она снималась и т. д.)

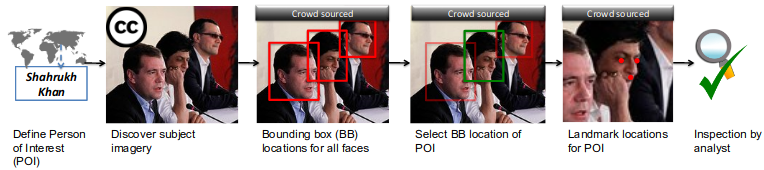
Около 30000 меток были тщательно проверены и выставлены экспертами вручную.

1. [VGGFace2](https://arxiv.org/pdf/1710.08092.pdf)

Состоит из 3.31 миллиона изображений 9131 знаменитостей, включающий в себя широкий спектр национальностей и профессий. Изображения различны по позам людей, их возрасту, по освещению и фону. Также набор сбалансирован по гендеру (58% мужчин). На каждого человека в среднем приходится 360 изображений ( от 80 до 843). Имеются подтверждённые людьми ограничительные рамки и пять опорных точек лица. Набор включает в себя тренировочных 8630 классов и 500 тестовых.



1. [IJB-A](https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Klare_Pushing_the_Frontiers_2015_CVPR_paper.pdf)

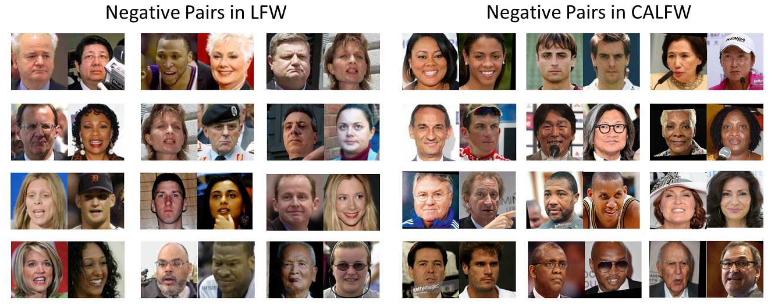
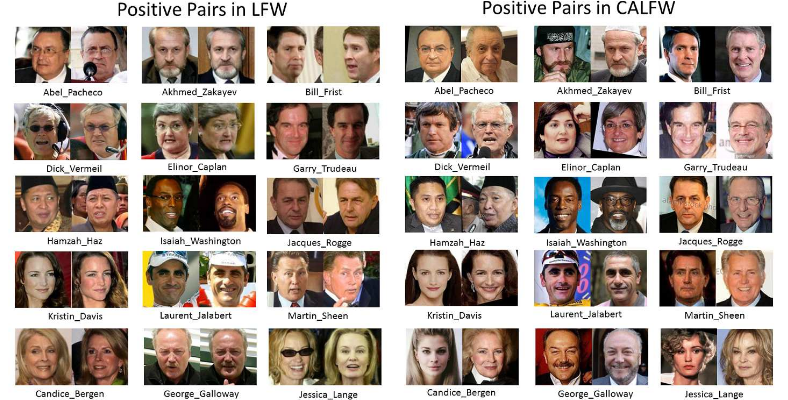
Состоит из 500 персон, изображенных в «естественной среде», вручную зафиксированы рамки (не только для интересующего лица, но и для всех лиц, попавших на изображение), а также опорные точки (центр глаза и нос, если они видны).

1. [Cross-Age LFW](https://arxiv.org/pdf/1708.08197.pdf)

Это модифицированный набор LFW с добавлением разницы в возрасте одного и того же человека.

Процесс построения:

1. сбор необработанных изображений из интернета;
2. запуск детектора лица и ручная проверка полученных результатов в случае, если на снимке присутствуют два и более человека;
3. обрезка и масштабирование обнаруженных лиц;
4. устранение дубликатов;
5. оценка корректности меток;
6. получение опорных точек лица и выравнивание;
7. оценка возраста и формирование тренировочного и тестового наборов: выбор пар с самой большой возрастной разницей в качестве положительных и людей с тем же полом и расой в качестве отрицательных пар.



1. [UTKFace](https://susanqq.github.io/UTKFace/)

Состоит из 20000 и более изображений в [естественной среде](https://drive.google.com/open?id=0BxYys69jI14kSVdWWllDMWhnN2c) или с [выровненными и обрезанными гранями](https://drive.google.com/drive/folders/0BxYys69jI14kU0I1YUQyY1ZDRUE?usp=sharing) разрешения 128x128 (присутствует только одно лицо) с 68 [опорными точками](https://drive.google.com/open?id=0BxYys69jI14kS1lmbW1jbkFHaW8) и метками о половой принадлежности ( 0 или 1 для мужского пола и женского), возрасте (от 0 до 106 лет) и этнической принадлежности (от 0 до 4 для людей с белой кожей, тёмной, азиатов и индийцев).



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Набор | Год | Изображений | Людей | Метки | «Чистота» меток | В естественной среде |
| AgeDB | 2017 | 16488 | 568 | Возраст с точностью до года | + | + |
| UmdFaces | 2017 | 367888 | 8277 | Информация об ограничительной рамке, принадлежности, выражении лица, гендере | + | - |
| MS-Celeb-1M | 2016 | ~10М | 100000 | Имя, связанная информация из открытой базы | + | - |
| VGGFace2 | 2018 | 3.31М | 9131 | Имя, рамки, опорные точки | + | + |
| IJB-A | 2015 | 5712 (+2085 видео) | 500 | Имя, опорные точки, рамки | + | + |
| CALFW | 2017 | 13233+ | 5749 | Имя, раса, гендер | + | + |
| UTKFace |  | 23708 |  | Имя, раса, гендер | + | + |