



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

SAMSUNG

НИУ ВШЭ, Научно-учебная групп «Анализ мультимедийных данных»  
Самсунг – ПОМИ им. Стеклова РАН «Центр исследования искусственного интеллекта»

# НЕЙРОСЕТЕВЫЕ ДЕТЕКТОРЫ В ЗАДАЧЕ АНАЛИЗА ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЯ ПО ФОТОГРАФИЯМ

ММРО 2019, Москва, 28 ноября 2019



---

# ПРЕДПОСЫЛКИ

- Развитие и распространение мобильных девайсов с возможностью фото- и видео-съёмки
- Взрывной рост количества медиа-информации пользователей
- Возможность извлечения информации о человеке-пользователе из его медиа
- Проблема безопасности и сохранности данных, приватности человеческой жизни



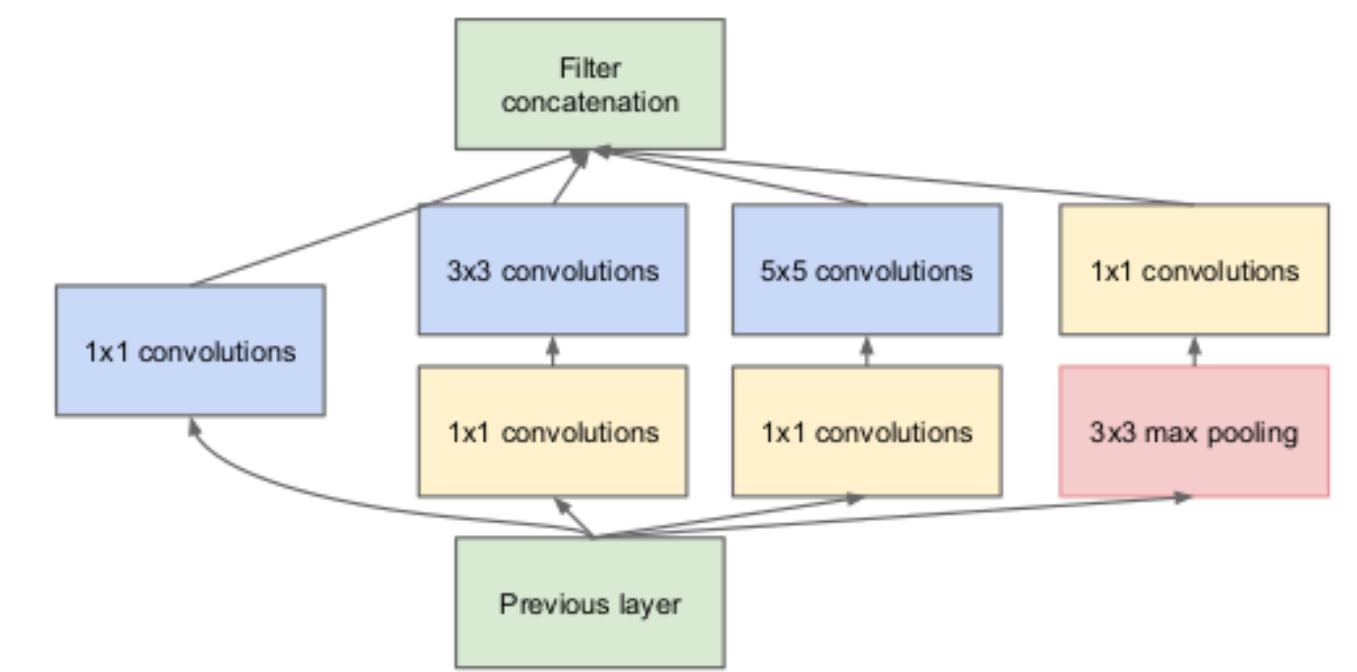
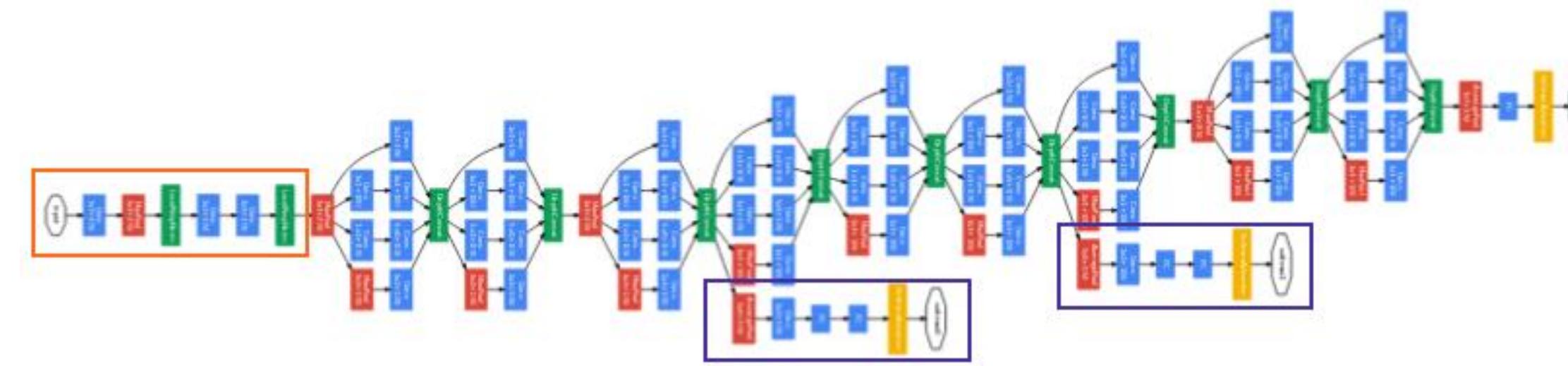
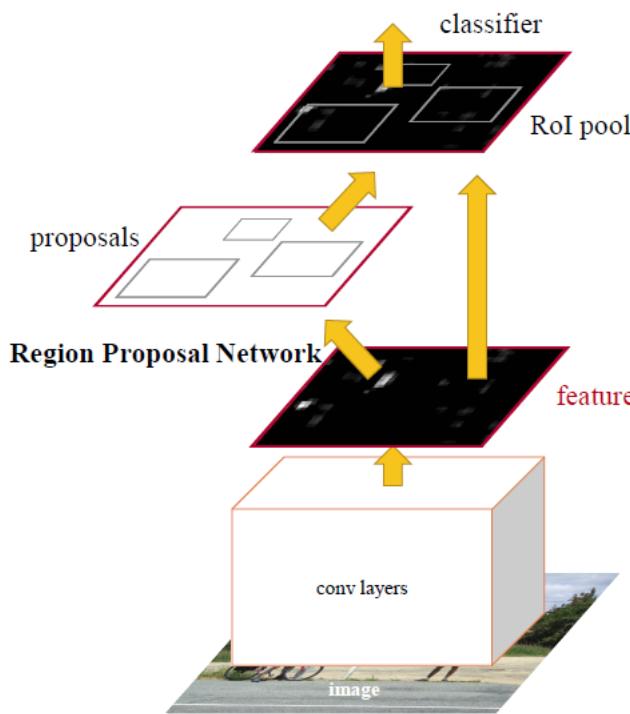
---

# ЦЕЛЬ И ЗАДАЧИ

- Цель: выяснить предпочтения пользователя на основе его медиа-информации
  - Предпочтения – некоторый вектор информации об интересах пользователя
  - Медиа-информация – содержимое галереи мобильного устройства (фото-видео)
- Задачи:
  - Обучить нейросетевые детекторы объектов на изображениях
  - Настроить определение приватной и публичной медиа-информации
  - Создать приложение для анализа галереи пользователя

# НЕЙРОСЕТЕВЫЕ АРХИТЕКТУРЫ

- Активно развивающаяся область
  - Классификаторы изображений: Inception, ResNet, MobileNet и др.
  - Детекторы объектов: Faster R-CNN, Single Shot Detector, YOLO.
- Возможность производить классификацию на удалённых серверах и на мобильных устройствах.

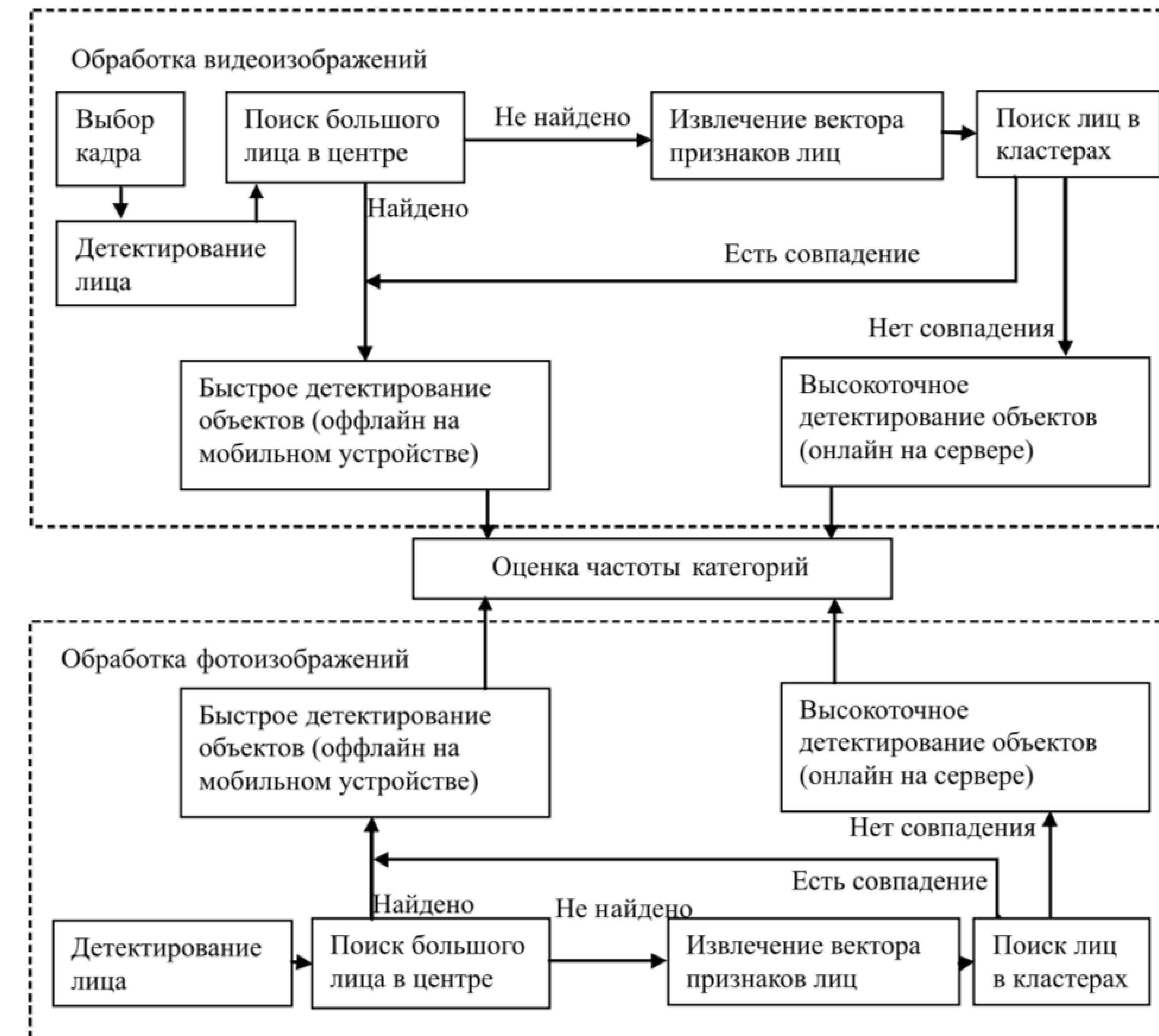




---

# ПРИВАТНОСТЬ ИНФОРМАЦИИ

- Многие фото/видео могут содержать чувствительную информацию
  - Такую информацию желательно никуда не передавать напрямую
- Решение: если на фото / видео есть лицо – это приватное фото / видео
  - Лицо может принадлежать владельцу / близкому человеку / другу
- Если фото / видео приватное – оно обрабатывается на устройстве пользователя, иначе – на сервере





# ОБУЧЕНИЕ МОДЕЛЕЙ

- Датасеты: OID v4, MS COCO, ImageNet
- Отобраны 146 категорий: строения, животные, транспорт, еда, и др.
- Для балансировки, каждый класс представлен не более чем 5000 объектов





# ОБУЧЕННЫЕ МОДЕЛИ

Детектор	Архитектура СНС (backbone)	Полнота	mAP	Полнота (родственные категории)	mAP (родственные категории)
Faster R-CNN	Inception ResNet v2	0,425	0,477	0,448	0,534
	Inception ResNet v2 (квантованная)	0,425	0,471	0,448	0,528
	Inception v3	0,393	0,537	0,414	0,593
	ResNet-50	0,332	0,583	0,35	0,636
	ResNet-101	0,465	0,562	0,485	0,618
SSDLite	MobileNet v2, 512 × 512	0,149	0,465	0,166	0,525
	MobileNet v2, 512 × 512 (квантованная)	0,149	0,463	0,166	0,524

- Родственные категории – близкие категории, определение которых нельзя считать ошибкой, например: канарейка-птица, кошка-животное.



# УЛУЧШЕНИЕ РЕЗУЛЬТАТОВ

Детектор	Архитектура СНС (backbone)	Число категорий	mAP (родственные категории)
Faster R-CNN	InceptionResNet v2	78	0,662
	InceptionResNet v2 (квантованная)	79	0,663
	Inception v3	44	0,762
	ResNet-50	30	0,838
	ResNet-101	67	0,76
SSDLite	MobileNet v2, $512 \times 512$	3	0,773
	MobileNet v2, $512 \times 512$ (квантованная)	3	0,768

- Отбор категорий под  $average\ recall > 0,75$
- Для некоторых моделей список категорий не включал важные для нас



# ПРИМЕРЫ





# ЭФФЕКТИВНОСТЬ МОДЕЛЕЙ

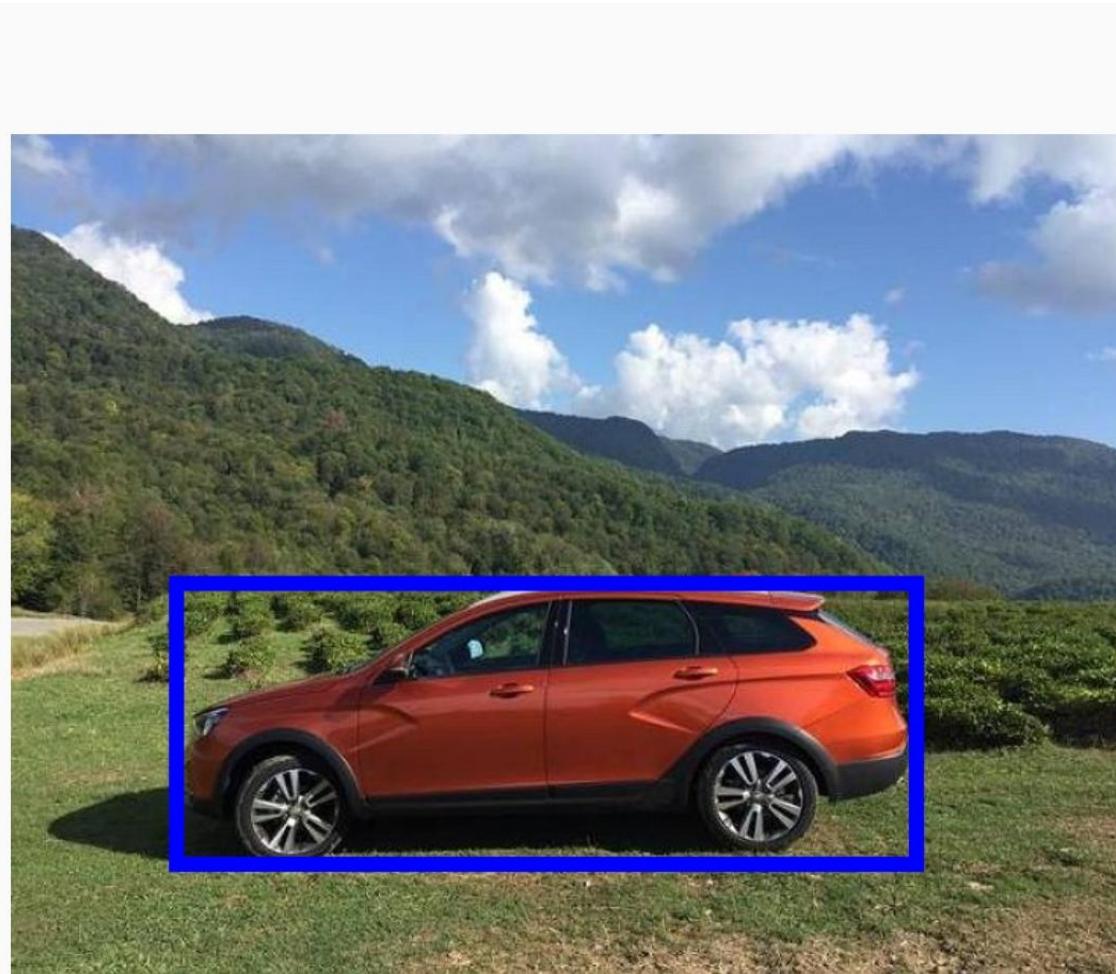
Детектор	Архитектура СНС (backbone)	Размер модели, Мбайт	Время детектирования, с	
			Ноутбук	Смартфон
SSDLite	MobileNet v2, 300 × 300	31,83	0,16	0,30
	MobileNet v2, 512 × 512	31,83	0,21	0,52
Faster R-CNN	Inception v2	64,91	0,4	1,25
	InceptionResNet v2	204,34	1,01	2,39



# АНАЛИЗ ДЕМОГРАФИИ

- Обучены отдельные модели для определения лиц
- Векторы признаков с предпоследнего слоя используются алгоритмами кластеризации для поиска часто встречающихся людей

Метод кластеризации	CHC	K/C	ARI	AMI	F-мера
Одиночная связь	VGGFace	4,10	0,440	0,419	0,616
	VGGFace2	3,21	0,580	0,544	0,707
	MobileNet	4,19	0,492	0,441	0,636
Метод невзвешенного попарного среднего	VGGFace	1,42	0,565	0,632	0,713
	VGGFace2	1,59	0,603	0,663	0,746
	MobileNet	1,59	0,609	0,658	0,751
Полная связь	VGGFace	0,95	0,376	0,553	0,595
	VGGFace2	1,44	0,392	0,570	0,641
	MobileNet	1,28	0,381	0,564	0,626
Метод взвешенного попарного среднего	VGGFace	1,20	0,464	0,597	0,662
	VGGFace2	1,05	0,536	0,656	0,710
	MobileNet	1,57	0,487	0,612	0,697
Медианное расстояние	VGGFace	5,30	0,309	0,307	0,516
	VGGFace2	4,20	0,412	0,422	0,742
	MobileNet	6,86	0,220	0,222	0,411
Ранговое расстояние	VGGFace	0,82	0,319	0,430	0,630
	VGGFace2	1,53	0,367	0,471	0,641
	MobileNet	1,26	0,379	0,483	0,652



PREV

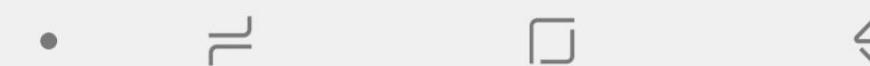
NEXT

photo 5 out of 753  
latitude=0,000 longitude=0,000

car (0,94)  
scenes:park/golf (0,97);  
No faces found  
text:

PREFERENCES

PHOTOS



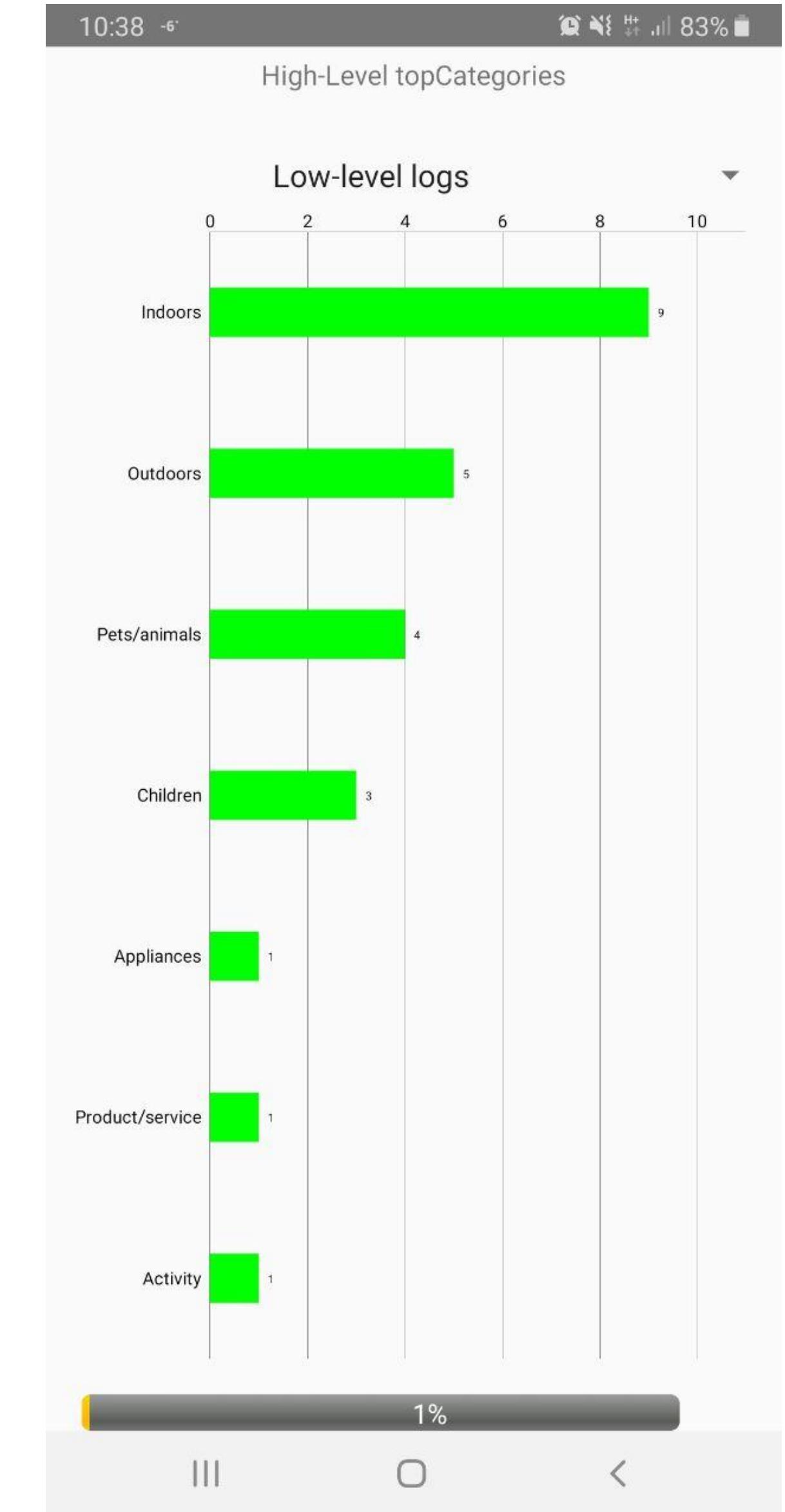
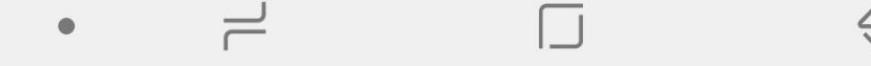
PREV

NEXT

photo 3 out of 753  
Detection: 3396 ms  
latitude=0,000 longitude=0,000

tower (0,55)  
tower (0,54)  
tower (0,52)  
tower (0,46)  
tower (0,40)  
tower (0,38)  
building (0,30)  
scenes:cinema (0,85);  
No faces found  
text:

PREFERENCES PHOTOS





# ИТОГИ

- Получены результаты по детектированию объектов на фотографиях пользователя
- Представлен метод разделения фотографий на публичные и приватные
- Результаты используются для определения предпочтений пользователя



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

E-mail: [igrechikhin@hse.ru](mailto:igrechikhin@hse.ru)