Роботу виконали студенти групи КА-71:

Чоловський Сергій, Іващенко Дмитро, Оксюта Ірина

**Вступ**

В основі нашого підходу до розпізнавання лежить ідея bag of visual words. Наш алгоритм – це комбо з дескриптора (ORB, SIFT або AKAZE), кластеризатора (K-Means) та класифікатора (варіації градієнтного бустингу над DescisionTree).

Для кластеризації використовувалась рандомізована версія алгоритму k середніх, реалізація – MiniBatchKMeans з sklearn.cluster.

Для класифікації використовувались реалізації GradientBoostingClassifier з sklearn.ensemble та XGBClassifier з xgboost.

Для кожного предмета будемо тренувати свій окремий класифікатор.

Етап навчання:

1. Здобуваємо з зображень дескриптори особливих точок, нормуємо отриману матрицю дескрипторів для кожного зображення за максимальним можливим значенням дескриптора.
2. Формуємо масив із дескрипторів усіх особливих точок з навчальної вибірки та навчаємо на цьому масиві кластеризатор – отримуємо об’єднання кластерів, де кожна особлива точка лежить у якомусь одному з них.
3. Для кожного зображення формуємо частотну характеристику – відношення кількості особливих точок зображення з кожного кластера до усіх особливих точок зображення.
4. Частотні характеристики усіх зображень з тренувальної вибірки з відповідними мітками класів «згодовуємо» класифікатору.

Етап роботи з «навченим» «комбо»:

1. Здобуваємо з зображень дескриптори особливих точок, нормуємо отриману матрицю дескрипторів для кожного зображення за максимальним значенням.
2. Для кожного зображення формуємо частотну характеристику – відношення кількості особливих точок зображення з кожного кластера до усіх особливих точок зображення.
3. За допомогою частотної характеристики та класифікатора отримуємо передбачення чи є об’єкт на зображенні.

**Дані для дослідження**

Об'єктами для класифікації були ті ж, що й у минулій лабораторній: кіт на ім'я Марсель, неваляшка та садовий гном, що схожий на Баліна.

Для кожного з об'єктів сформовано свій набір фотографій. Для кожної локації робилися окремі фото де об'єкт є, та його немає, також додавалися інші фото без об'єкта. Фото робилися за різного освітлення та з різних ракурсів для кожної локації.

Фото для навчання класифікатора неваляшок формувалися поступово, з урахуванням впливу кожного фото на результати навчання, найбільш «шкідливі» фото у кінцеву вибірку не потрапили, також у вибірці є декілька фото загальних планів без неваляшки. Усі фото зняті на цифрову камеру 64 МП й з роздільною здатністю 3120 на 4160 пікселів.

Вибірка для кота складається з типових позицій кота у його середовищі проживання, та фото його середовища, також є деяка кількість інших фото в якості негативних прикладів. Фото кота зроблені у двох режимах: високоякісні фото з високою роздільною здатністю(3120 на 4160 пікселів) та менш якісні у HD. На деяких фото Марсель зображений великим планом(його морда), на деяких його видно цілком.

Вибірка для гнома складається з власне гнома у різних місцях, та низки фото з інших двох вибірок. Тобто, ця вибірка найрізноматніша у плані негативних прикладів.

Остаточна перевірка даних здійснювалася на наборах зображень, що не потрапили у наші вибірки та з доступних у мережі інтернет фото різних об’єктів: котів, собак, квітів, матрьошок, облич людей, фруктів, різноманітного транспорту та ін.

Також записано та змонтовано невелике відео у якому присутні усі досліджувані об’єкти, на основі результатів роботи класифікаторів створено три відео: по одному для дескрипторів SIFT,ORB та AKAZE.

**Вибір параметрів моделі**

Градієнтний бустинг має наступні параметри: максимальна глибина дерева, кількість дерев, learning rate. Також ми можемо обирати різну кількість кластерів. Кількість кластерів виступає в ролі гіперпараметра. Процедура вибору виконується для кожного дескриптора окремо.

Побудова системи розпізнавання неваляшки

Градієнтний бустинг має наступні параметри: максимальна глибина дерева, кількість дерев, learning rate. Також ми можемо обирати різну кількість кластерів. Зокрема, для навчання на даних неваляшки було використано наступну стратегію:

1. Параметри класифікатора шукались за допомогою крос-валідації на тренувальній вибірці за фіксованою кількістю кластерів за середніми значеннями метрик Accuracy та Balanced Accuracy (Accuracy з урахуванням розмірів класів) по розбиттям.

|  |  |  |  |
| --- | --- | --- | --- |
| Параметр | Мінімум | Максимум | Крок |
| Кількість дерев | 10 | 80 | 1 |
| Максимальна глибина дерева | 2 | 5 | 1 |
| Learning rate | 0.001 | 0.5 | Проміжні значення: 0.005, 0.01, 0.05, 0.1, 0.5 |

1. У ролі гіперпараметра виступала кількість кластерів, тож цей параметр обирався на даних з відкладеної вибірки за метриками Accuracy, Precision та Recall. Було обрано декілька найкращих моделей за кількістю кластерів, надалі працювали з ними.

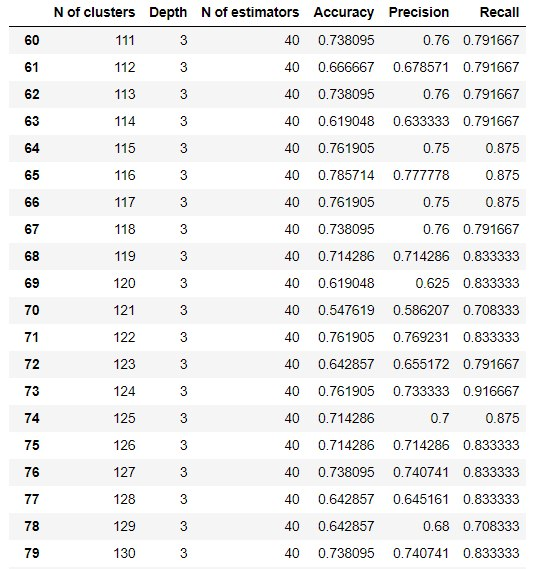


Рис. 1 Приклад таблиці з результатами метрик на відкладеній вибірці, за якою обиралися моделі (у даному випадку для дескриптора AKAZE)

Таким чином, обрано наступні моделі:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Дескриптор | Кількість кластерів | Кількість дерев | Максимальна глибина дерева | Learning rate |
| SIFT | 132 | 16 | 4 | 0.05 |
| AKAZE | 97 | 10 | 3 | 0.5 |
| ORB | 172 | 40 | 3 | 0.05 |

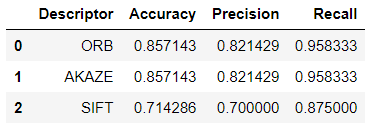


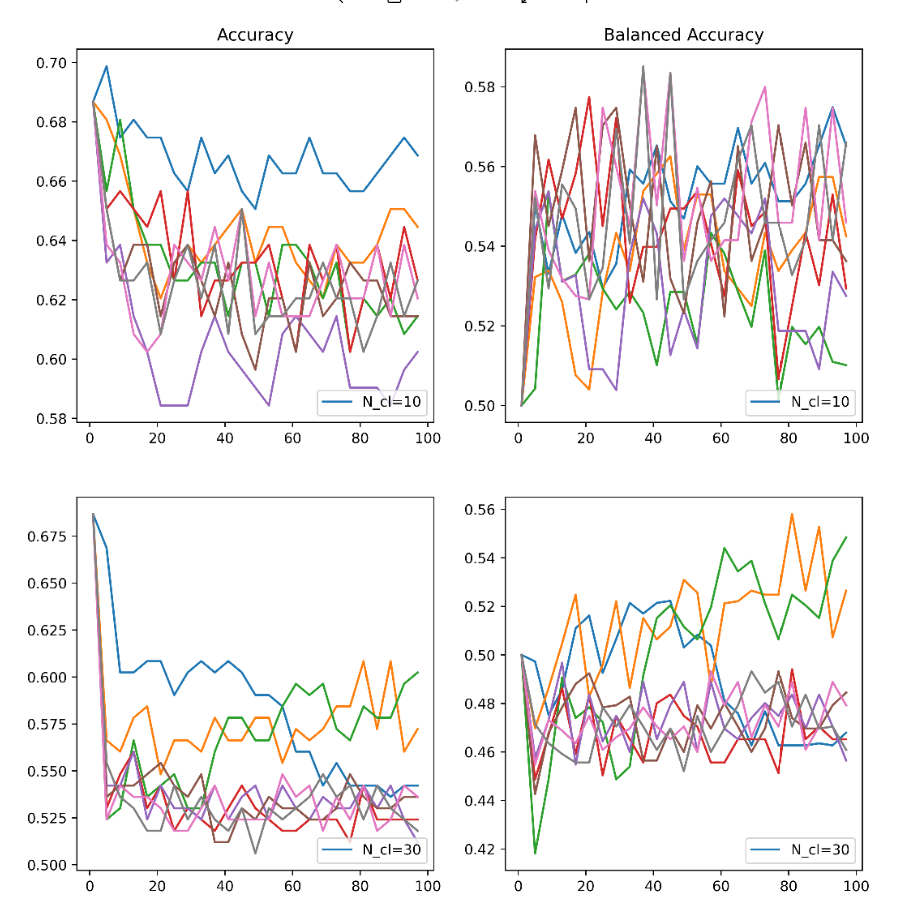
Рис. 2 Результати метрик на тестовій вибірці

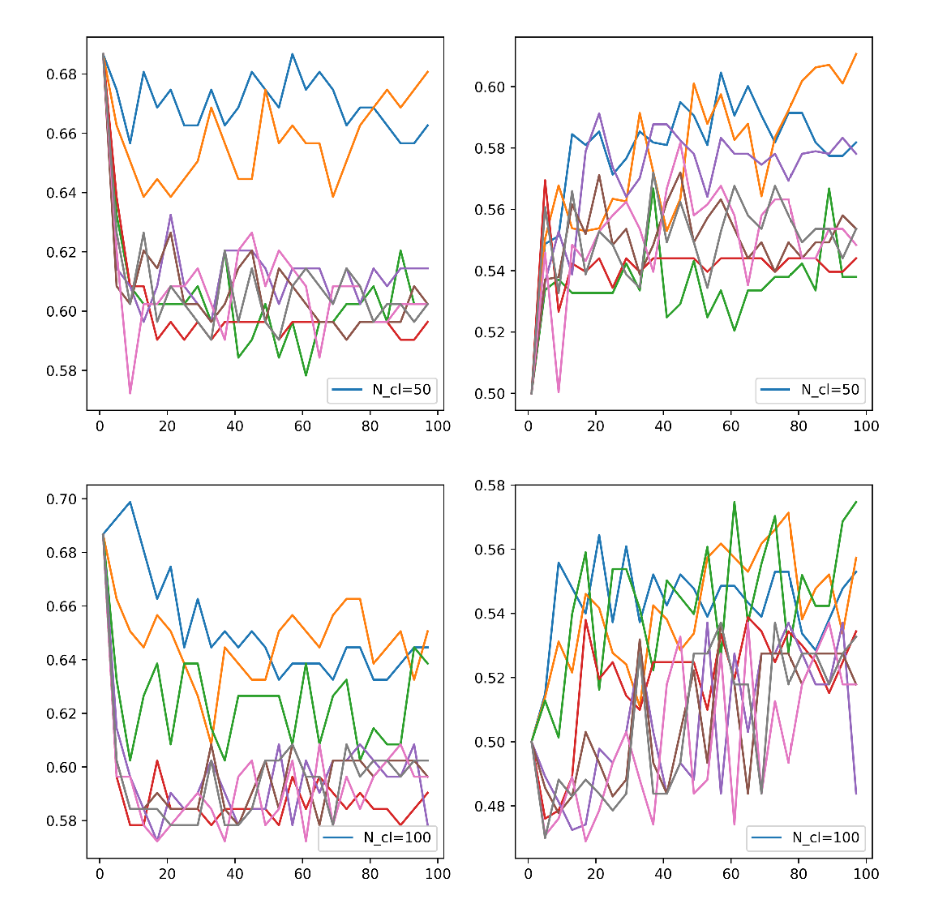
Вибір параметрів класифікатора для кота

1. Параметри класифікатора шукались за допомогою крос-валідації на тренувальній вибірці за фіксованою кількістю кластерів за середніми значеннями метрик Accuracy та Balanced Accuracy (Accuracy з урахуванням розмірів класів) по розбиттям. При виборі параметрів Balanced Accuracy має перевагу.
2. У ролі гіперпараметра виступала кількість кластерів, тож цей параметр обирався на даних з відкладеної вибірки за метриками Accuracy, Precision та Recall.

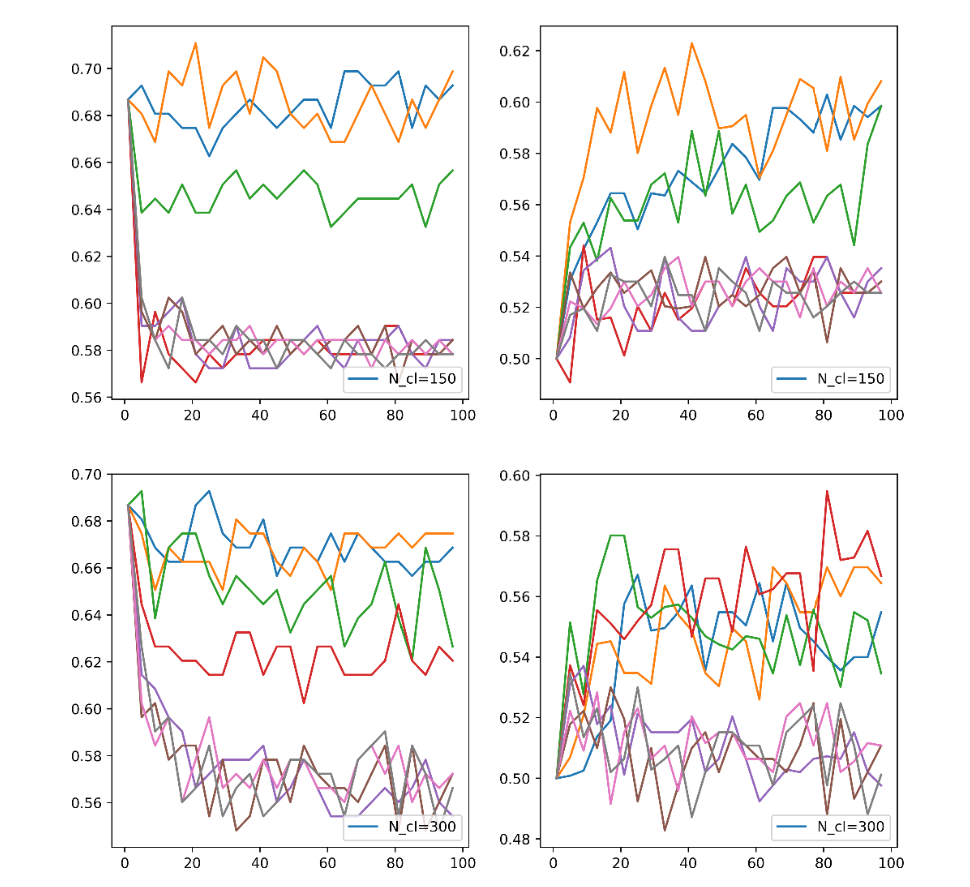
|  |  |  |  |
| --- | --- | --- | --- |
| Параметр | Мінімум | Максимум | Крок |
| Кількість дерев | 1 | 100 | 4 |
| Максимальна глибина дерева | 2 | 9 | 1 |
| Кількість кластерів | Можливі значення:   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 10 | 30 | 50 | 100 | 150 | 300 | 500 | | | |

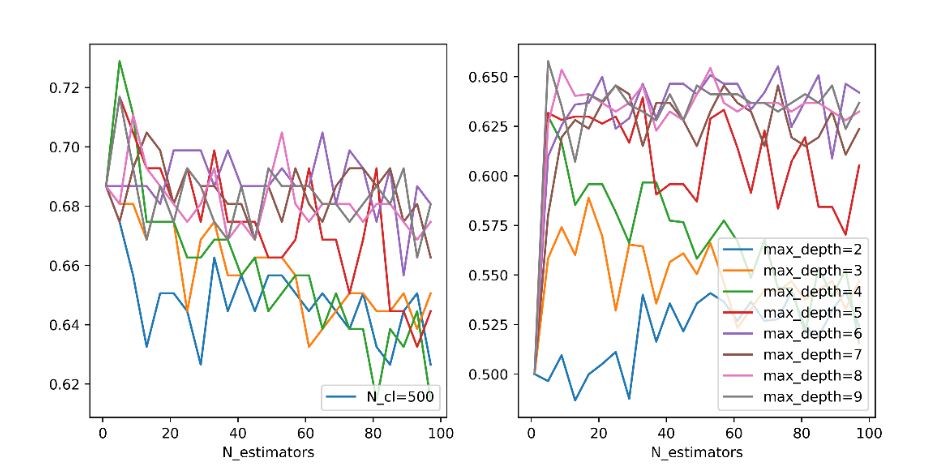
В результаті обчислень отримано наступні криві навчання:

 Рис. 3 криві навчання для різної так різного числа кластерів для дескриптора AKAZE



Продовження Рис. 3





Кінець Рис. 3

Як бачимо на рис(3) для усіх моделей точність має локальний максимум на початку. Те ж помітно й на кривих навчання для інших дескрипторів. Також бачимо, що моделі показують набагато кращу динаміку, але все одно приблизно після 50 дерев у моделі навчання зупиняється. Можемо припустити, що наші дані насправді мають досить просту статистичну структуру.

На таблицях нижче наведено результати найкращих моделей за крос-валідацією.

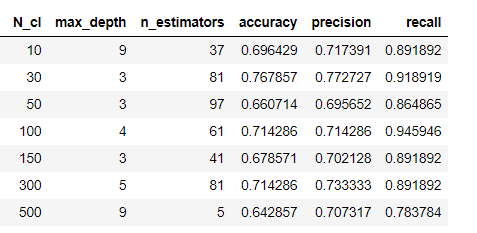


Рис. 4 Таблиця з метриками з параметрами для кожного дескриптора та моделі

Для дескриптора AKAZE

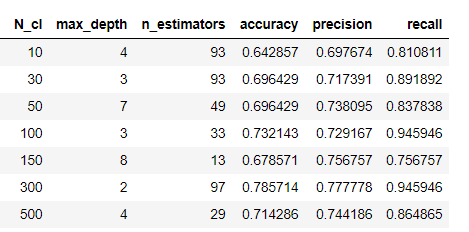


Рис. 5 Таблиця з метриками з параметрами для кожного дескриптора та моделі

Для дескриптора SIFT

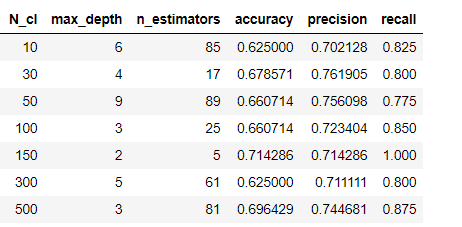


Рис. 6 Таблиця з метриками з параметрами для кожного дескриптора та моделі

Для дескриптора ORB

Ця процедура виконувалась для кожного дескриптора.

Таким чином, обрано наступні моделі:

|  |  |  |  |
| --- | --- | --- | --- |
| Дескриптор | Кількість кластерів | Максимальна глибина дерева | Кількість дерев |
| SIFT | 300 | 2 | 97 |
| SIFT | 100 | 3 | 33 |
| ORB | 150 | 2 | 5 |
| AKAZE | 30 | 3 | 81 |
| AKAZE | 100 | 4 | 61 |

Для дескрипторів SIFT та AKAZE навіть за метриками та кривими навчання не вдалося обрати один найкращий дескриптор.

Пошук оптимальних параметрів класифікатора для гнома

1. За допомогою крос-валідації на тренувальній вибірці за фіксованою кількістю кластерів за середніми значеннями метрик Accuracy та Balanced Accuracy (Accuracy з урахуванням розмірів класів) по розбиттям було обрано параметри класифікації. При виборі параметрів Balanced Accuracy має перевагу.

|  |  |  |  |
| --- | --- | --- | --- |
| Параметр | Мінімум | Максимум | Крок |
| Кількість дерев | 1 | 100 | 4 |
| Максимальна глибина дерева | 2 | 9 | 1 |
| Кількість кластерів | Можливі значення:   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 10 | 30 | 50 | 100 | 150 | 300 | 500 | | | |

1. У ролі гіперпараметра також виступала кількість кластерів, цей параметр обирався на даних з відкладеної вибірки за метриками Accuracy, Precision та Recall.

В таблицях нижче наведено результати моделей за крос-валідацією.

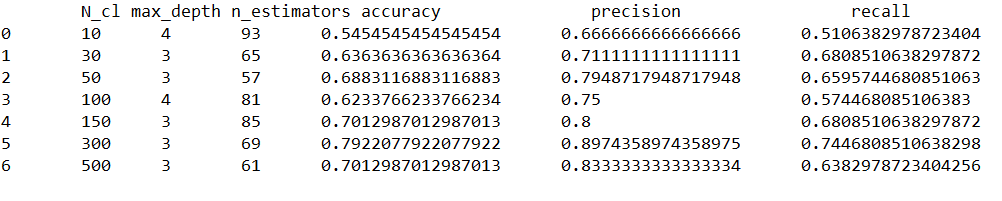


Рис. 7 Таблиця з метриками та параметрами моделей для дескриптора SIFT

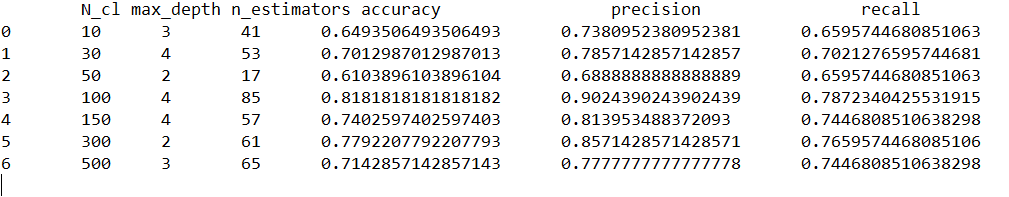
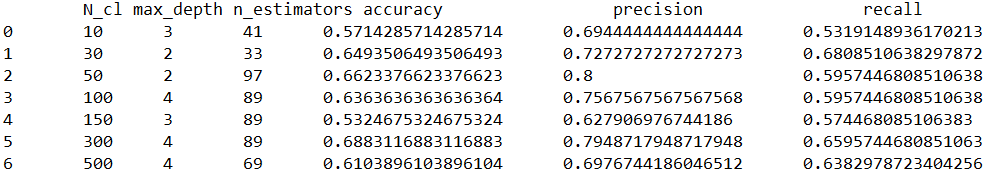
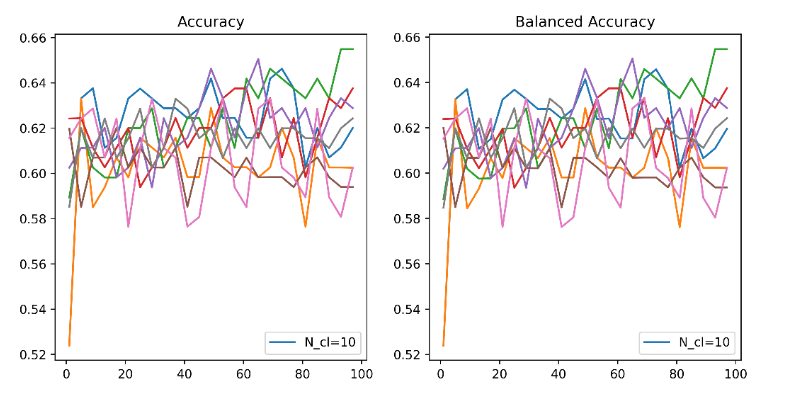


Рис. 8 Таблиця з метриками та параметрами моделей для дескриптора AKAZE

Рис. 9 Таблиця з метриками та параметрами моделей для дескриптора ORB

Графіки точності моделей в залежності від максимальної глибини дерев:



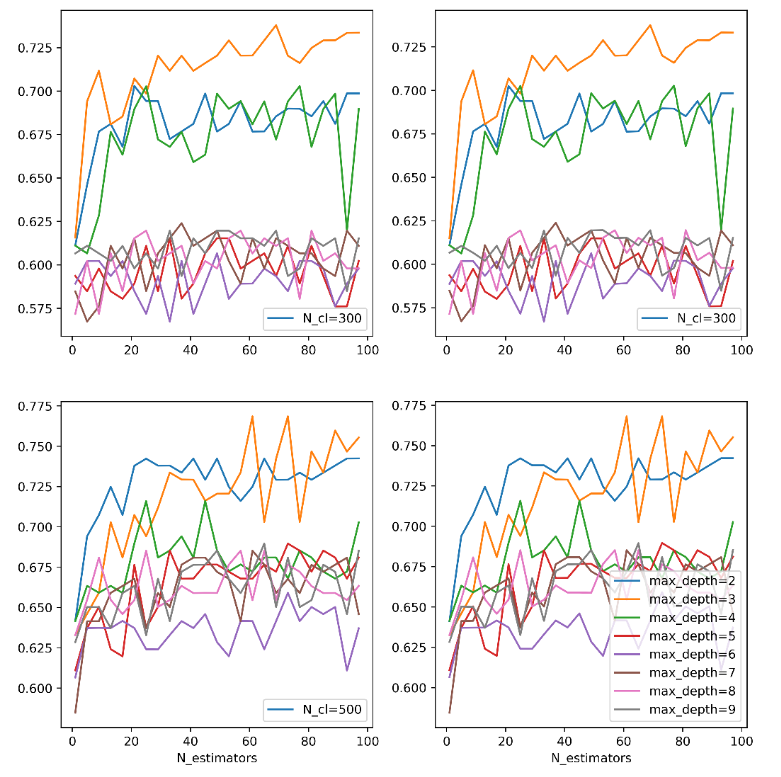
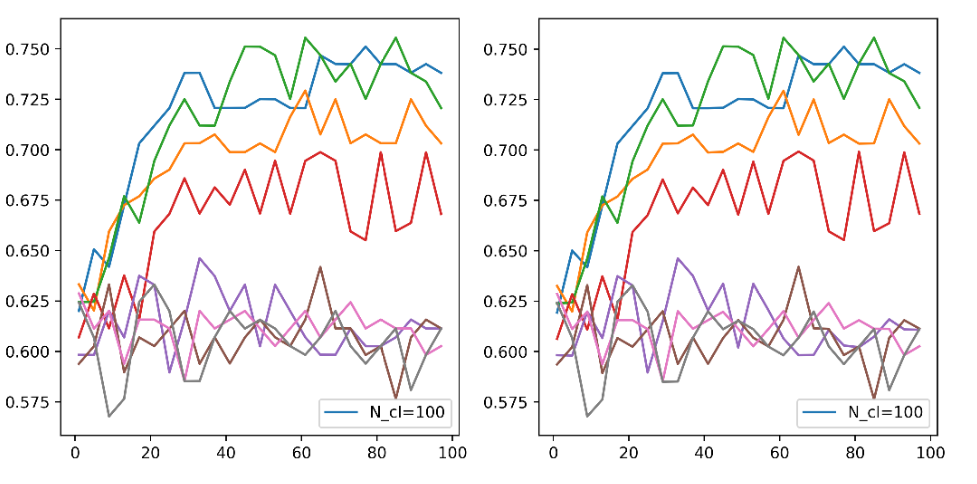
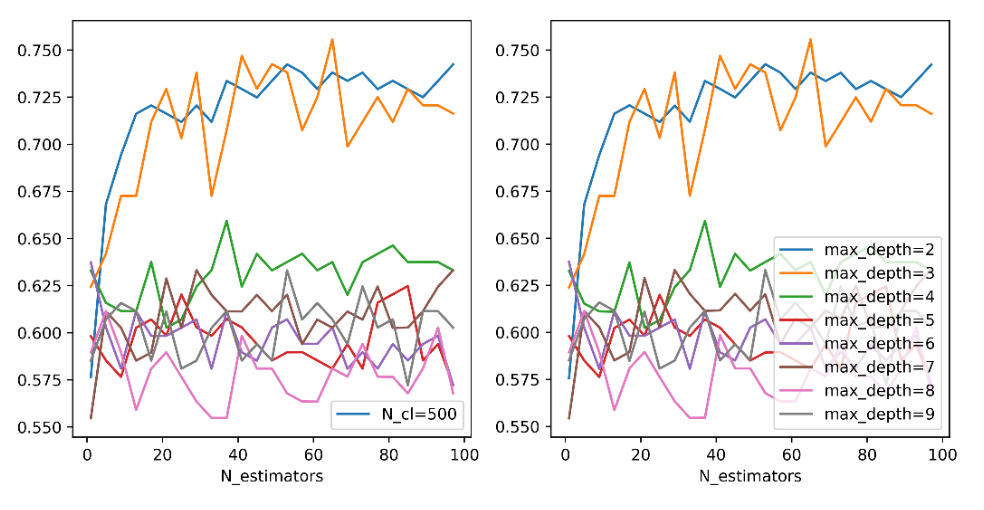
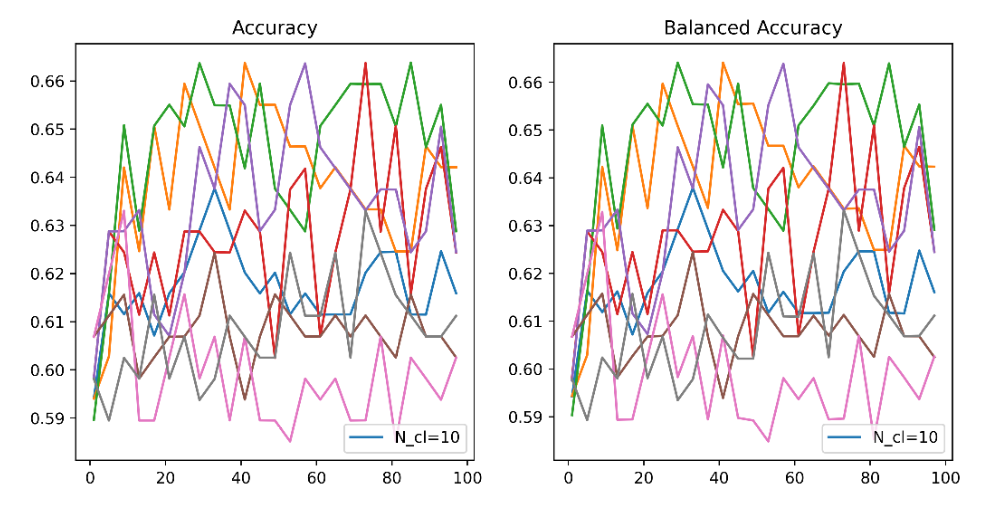
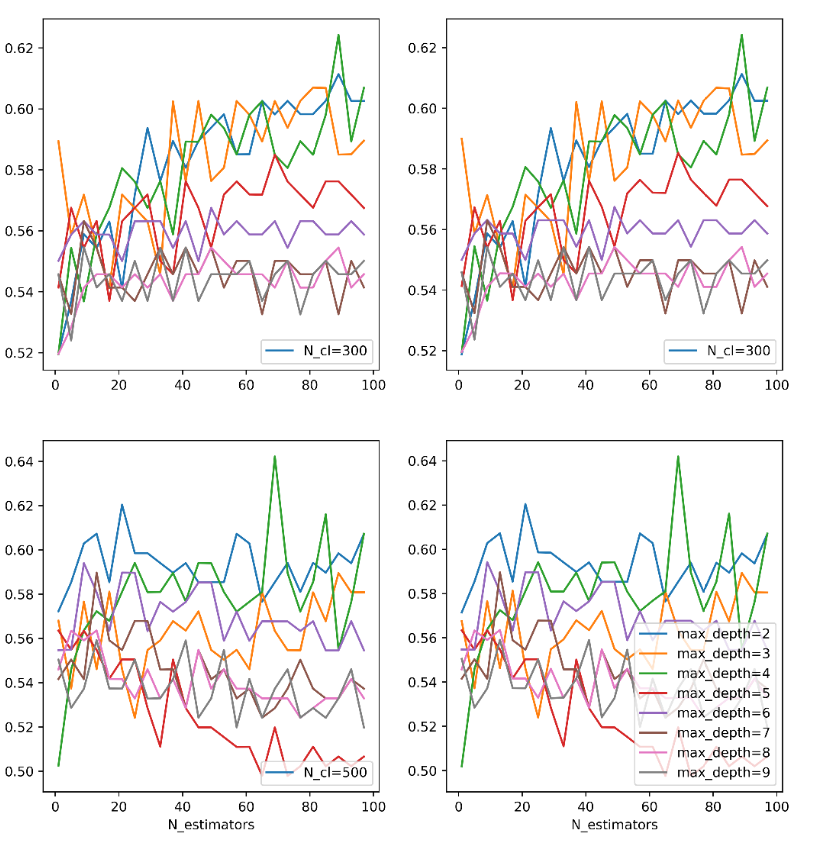
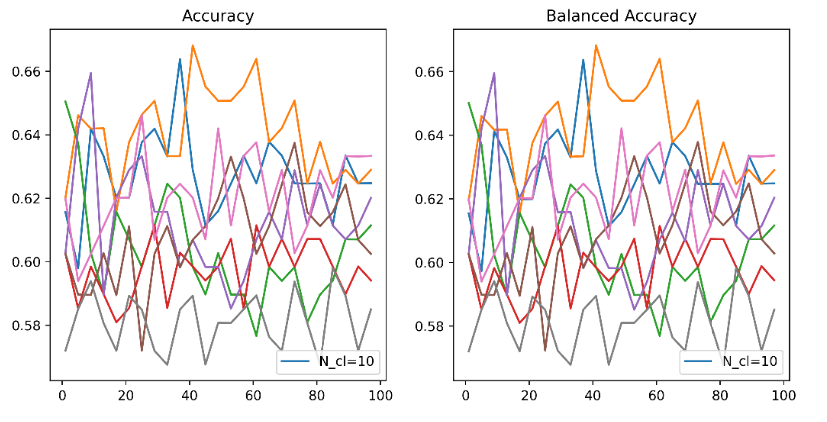


Рис. 10 Графіки з метриками точності моделей для дескриптора SIFT

 Рис. 12 Графіки з метриками точності моделей для дескриптора AKAZE

 Рис. 13 Графіки з метриками точності моделей для дескриптора ORB

В результаті обираємо наступні моделі:

|  |  |  |  |
| --- | --- | --- | --- |
| Дескриптор | Кількість кластерів | Кількість дерев | Максимальна глибина дерева |
| SIFT | 300 | 69 | 3 |
| AKAZE | 100 | 85 | 4 |
| ORB | 300 | 89 | 4 |

Результуючі метрики

Виміряно час навчання й роботи найкращих моделей на своїх навчальних та відкладених, результати наведені в таблиці.

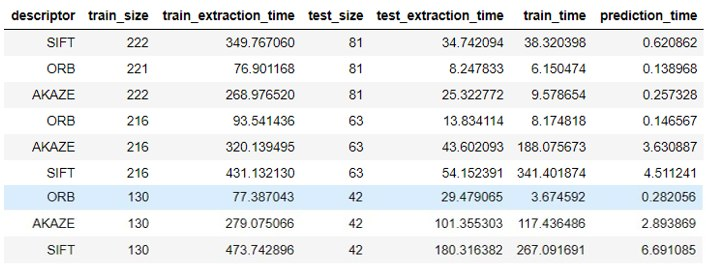


Рис.14 Таблиця витрат часу.

Результати витрат часу відповідають попередній лабораторній та теоретичним відомостям. Також при створенні демонстраційних відео ми міряли середній час обробки кадру.

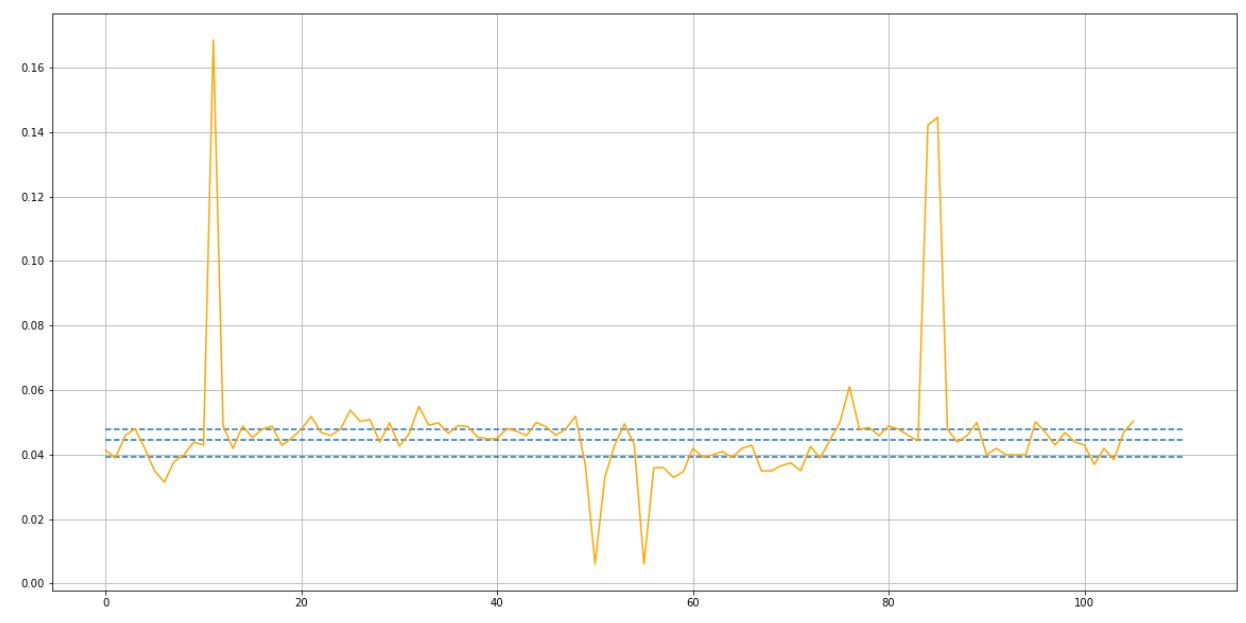
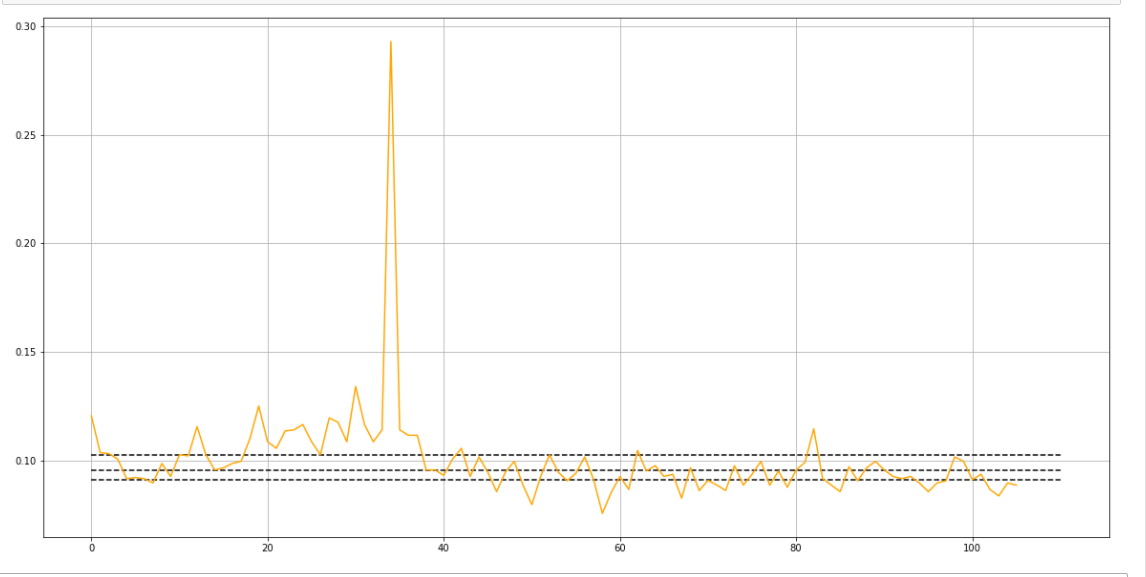
Рис. 15 Час обробки кадрів моделлю ORB, з квартилями

Рис. 16 Час обробки кадрів моделлю AKAZE, з квартилями

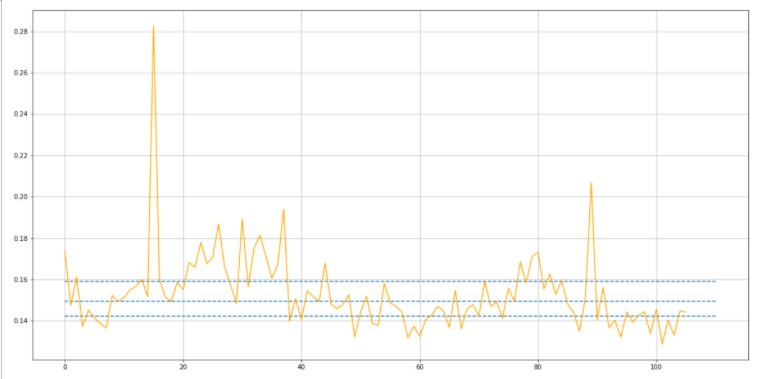


Рис. 17 Час обробки кадрів моделлю SIFT, з квартилями

Як бачимо ORB цілком придатний для виконання онлайн, AKAZE – також можна застосовувати онлайн, але слід уважно слідкувати за часом обробки кожного окремого кадру. При роботі з SIFT доведеться не розпізнавати щонайменше кожен 10й кадр, тож онлайн застосування в умовах наступної лабораторної сумнівне.

Також виміряно обчислено показники роботи усіх алгоритмів на тестових об’єктах. Отримані результати роботи можна знайти у файлах test.xslsx та test.csv

Так, ORB-класифікатор кота практично завжди каже, що кіт є, що свідчить про непридатність отриманого класифікатора для практичного застосування.

Всі класифікатори кота досить добре розпізнали котів у відповідному датасеті, досить посередньо впоралися з транспортними засобами та людьми, але повністю провалилися з неваляшками та матрьошками. Також вони часто плутали котів та собак. Фрукти розпізнавалися як коти досить рідко. По парам класифікаторів у дескрипторах приблизний паритет. Ми схильні вважати, що дещо краще впорався SIFT на 300 кластерів.

Усі класифікатори гномів впоралися добре, а з урахуванням складності моделей – то дуже добре. Можливо це пояснюється різноманітністю вибірки, а можливо – особливостями самого гнома( у нього дуже багато особливих точок на ковпаку).

Класифікатори неваляшок розпізнали гномів, але з іншими задачами, в середньому, впорались навіть гірше за котів. На нашу думку, причиною цього може бути overfitting на досить бідній вибірці.

За спостереженнями по якості результатів найкраще у середньому показує себе AKAZE, дескриптор ORB показує себе дещо гірше, ніж SIFT та AKAZE. Для неваляшки дескриптор SIFT показав себе гірше, можливо, це пов’язано з особливостями, що описані вище.

**Висновки**

В ході виконання роботи підготовано навчальні й тестові дані, отримано досвід навчання примітивної моделі на обмеженій вибірці. Високих показників точності та низьких ймовірностей помилок першого та другого роду досягти не вдалося. Однак, в цілому, підхід Bag of visual words виявися цілком придатним до застосування на вибірках не дуже різноманітних, у певному сенсі, зображень. Деякі з отриманих нами моделей виявились перенавченими й зовсім непридатними до застосування, а деякі досить стабільними. Для обгрунтованої відповіді на питання яка кількість кластерів краща у подібних задачах нам не вистачає даних. Наші ж спостереження не показують суттєвої різниці на зображеннях, що не схожі на тренувальні. На осноі аналізу кривих навчання також висунено припущення про можливість розв’язання нашої задачі деяким примітивним класифікатором( персептроном, одиничним деревом).

Слід зазначити, що якість записаного нами відео може бути вищою, але в ході суміщення результатів роботи виконавців виявилось, що в ході роботи використовувались різні версії бібліотек, що не могло позитивно вплинути на результати обробки відео( файл video.ipynb).

Враховуючи швидкодію та якість роботи, можемо сказати, що для виконання наступної роботи більш підходящою буде модель заснована на AKAZE. Можливо, якщо ми вдосконалимо навчальні вибірки ORB покаже прийнятну якість. SIFT через свою повільність однозначно непридатний для онлайн застосування.