

СРАВНИТЕЛЬНЫЙ АНАЛИЗ НАБОРОВ ДАННЫХ ДЛЯ ОЦЕНКИ КАЧЕСТВА ВИДЕО И ИЗОБРАЖЕНИЙ

Борисовский Лев Юрьевич

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: levbor888@graphics.cs.msu.ru

Научный руководитель — Ватолин Дмитрий Сергеевич

Задача анализа объективного качества видео или изображений заключается в разработке метрики, на вход которой подается видео или изображение, а на выходе получается число, количественно описывающее его качество. Это число должно согласовываться с оценкой, полученной от людей, - субъективным качеством.

Оценки от людей получаются с помощью экспериментов, в ходе которых участники оценивают качество показываемых материалов, используя числовые шкалы. Оценка (MOS, Mean Opinion Score) для каждого объекта из набора данных - результат усреднения полученных оценок. Такие наборы данных называются субъективными. С помощью субъективных наборов данных разрабатываются модели, как правило, машинного обучения, нацеленные получить для объекта оценку его качества.

В последние годы развитие нейросетевых подходов существенно способствовало появлению различных методов решения этой задачи. Однако, нейросетевые подходы сильно привязаны к качеству и объему обучающих выборок. Из-за низкого качества разметки и разнообразия примеров обученные нейросетевые модели оказываются неприспособленными к видео и изображениям, возникающим в реальных задачах.

Все чаще авторы субъективных наборов данных прибегают к краудсорсингу [1]. Этот подход обладает гораздо меньшей устойчивостью к человеческим ошибкам. Использование краудсорсинга затрудняет оценку адекватности условий просмотра материалов субъектами, а также их добросовестность.

В данной работе предлагается набор методов, позволяющий оценить качество субъективных наборов данных по таким критериям как разнообразие, полнота искажений и качество разметки. В результате анализа разработана метрика согласованности субъективной разметки с реальным качеством представленных в наборе объектов. С помощью данного подхода составлен рейтинг 25 открытых субъективных наборов данных.

В первую очередь анализируется разнообразие видео и изображений. Для этого, во-первых, используется классический подход [2], основанный на подсчете низкоуровневых статистик: SI (Spatial Information) - энергия границ, CF (Colorfulness) - красочность и TI (Temporal Information) - информация о движении. Далее на основе этих статистик считаются различные меры изменчивости и разнообразия.

Недостатком такого подхода является слишком низкий уровень абстракции, не позволяющий в полной мере оценить разнообразие с точки зрения человека. Поэтому в данной работе предлагается использовать извлеченные из видео и изображений нейросетевые признаки, полученные с помощью моделей классификации (напр. ResNet [3]). Эти признаки наиболее точно выражают впечатление от просмотра с точки зрения человека. Затем на основе этих признаков рассчитываются меры изменчивости и метрики генеративных моделей компьютерного зрения, такие как Vendi Score [4].

После этого анализируется качество представленной разметки - для этого проводится трехэтапный поиск выбросов.

Первый этап: подсчет лучших метрик оценки объективного качества.

Второй этап: использование полученных значений в качестве признакового пространства и поиска в нем условных относительно MOS выбросов. Здесь используется предположение о том, что неточная оценка MOS будет сильно расходиться с ансамблем лучших метрик качества. Инструментом поиска выбросов выступают алгоритмы машинного обучения [5].

Третий этап: построение приложения (Рис. 1) для ручного подтверждения выбросов. Для этого специальным алгоритмом собираются пары видео или изображений, которые близки по MOS, но расходятся по метрикам. После просмотра таких пар удастся определить, кто ошибается - разметка или метрики качества.

В ходе лабораторного исследования для каждого найденного примера валидируется отнесение его к категории выбросов.

Мера согласованности субъективной разметки с реальным качеством объектов из набора данных вычисляется как взвешенная на количества подтверждающих пар доля найденных алгоритмом выбросов.

Итогом анализа является совокупный индекс общей пригодности субъективного набора данных к обучению на нем моделей оценки качества.

Иллюстрации

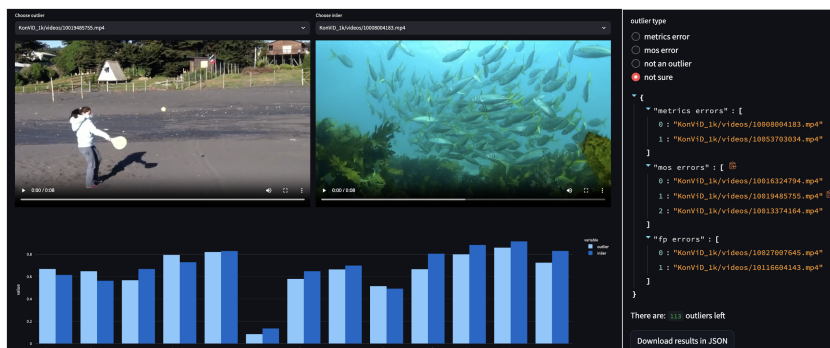


Рис 1. Вид приложения для подтверждения выбросов в разметке

Литература

1. Götz-Hahn F. et al. KonVid-150k: A dataset for no-reference video quality assessment of videos in-the-wild //IEEE Access. – 2021. – Т. 9. – С. 72139-72160.
2. Winkler S. Analysis of public image and video databases for quality assessment //IEEE Journal of Selected Topics in Signal Processing. – 2012. – Т. 6. – №. 6. – С. 616-625.
3. He K. et al. Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 770-778.
4. Friedman D., Dieng A. B. The vendi score: A diversity evaluation metric for machine learning //arXiv preprint arXiv:2210.02410. – 2022.
5. Ruff L. et al. A unifying review of deep and shallow anomaly detection //Proceedings of the IEEE. – 2021. – Т. 109. – №. 5. – С. 756-795.