

РАЗРАБОТКА АДАПТИВНОГО МЕТОДА РАНЖИРОВАНИЯ МОДЕЛЕЙ ПРЕДСКАЗАНИЯ КАРТ ВНИМАНИЯ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ

Брынцев Алексей Юрьевич

Студент

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

E-mail: alexey.bryncev@graphics.cs.msu.ru

Научный руководитель — Ватолин Дмитрий Сергеевич

Карта внимания (салиентности) кадра видео представляет из себя одноканальную тепловую карту, на которой более яркие области соответствуют областям кадра с более высоким значением визуального внимания. Обычно такие карты получают при помощи специального устройства для отслеживания глаз — eye-tracker. Задача предсказания карт салиентности заключается в автоматическом получении этих карт без задействования наблюдателей и дополнительных дорогостоящих устройств.

Развитие моделей глубокого обучения в последние годы повлекло за собой значительное увеличение количества различных нейросетевых алгоритмов решения задачи предсказания карт внимания. Поэтому методы сравнения качества работы этих алгоритмов стали актуальными, и начали выкладываться результаты сравнений (бенчмарки), например, DNF1K [1].

Данные, собирающиеся из eye-tracker, представляют из себя набор точек просмотра на кадре с привязкой ко времени. Такие данные собираются от нескольких наблюдателей и потом обрабатываются вместе. Обработка включает в себя такие процедуры, как устранение саккад и размытие точек фильтром Гаусса для аппроксимации эталонных данных. К сожалению, авторы разных датасетов используют разные параметры фильтра Гаусса при генерации данных. Модель с различными параметрами обучающей и тестовой выборки будет выдавать результаты хуже, чем при тестировании на выборке с теми же параметрами. С этой точки зрения использование исходных результатов моделей при сравнении, как в DNF1K [1], нельзя назвать в достаточной степени объективным.

В данной работе используется метод доменной адаптации, который устраняет вышеуказанную проблему. Описание метода представлено в работе SAVAM [2]. Суть метода заключается в дополнительной постобработке результатов двумя функциями: монотонным нелинейным преобразованием яркости и подмешивании с коэффици-

ентом карты внимания Center Prior (карта saliентности полученная единственной точкой по центру кадра). Для всех результатов каждой модели решается задача оптимизации этого функционала по 256 коэффициентам преобразования яркости и 1 коэффициенту подмешивания. Для использования производительного оптимизатора задача приводится к следующему виду:

$$(\beta, \mathbf{m}) = \underset{\substack{\mathbf{x}_1 > 0 \\ 0 < \mathbf{x}_i < \mathbf{x}_{i+1}, i > 1}}{\operatorname{argmin}} \frac{1}{2} \mathbf{x}^T \mathbf{H} \mathbf{x} + \mathbf{f}^T \mathbf{x} + c,$$

где $\mathbf{x} = (\beta, \mathbf{m}) \in \mathbb{R}^{N+1}$ содержит целевые параметры; матрица $\mathbf{H} \in \mathbb{R}^{(N+1) \times (N+1)}$, вектор $\mathbf{f} \in \mathbb{R}^{N+1}$ и число c определяют задачу оптимизации; а \mathbf{H} — это Эрмитова матрица.

В качестве датасета использовался многотиповой датасет высокого разрешения SAVAM [2]. Данные собирались при помощи 500 Гц eye-tracker с участием 50 наблюдателей.

Ниже представлена малая часть результатов сравнения до использования доменной адаптации и после (сортировка по SIM):

До использования Доменной Адаптации						
Модель	SIM	CC	NSS	AUC Judd	KLDiv	FPS
ViNet	0.613	0.696	2.35	0.855	0.583	1.10
ACLNet	0.579	0.615	1.87	0.834	1.529	4.18
MSI-Net	0.571	0.631	1.94	0.827	2.134	1.28
После использования Доменной Адаптации						
Модель	SIM	CC	NSS	AUC Judd	KLDiv	FPS
ViNet	0.627	0.733	2.13	0.864	0.497	1.10
HD2S	0.615	0.707	1.89	0.844	0.545	24.51
TASED-Net	0.610	0.710	1.96	0.852	0.538	1.85

Литература

1. Wang W. Revisiting Video Saliency: A Large-scale Benchmark and a New Model, // In IEEE Computer Vision and Pattern Recognition Conference, 2018.
2. Lyudvichenko V. A semiautomatic saliency model and its application to video compression // In IEEE International Conference on Intelligent Computer Communication and Processing, 2017, P. 403–410.