УДК 004.89

ИСПОЛЬЗОВАНИЕ ВЕЙВЛЕТОВ И МЕХАНИЗМА ВНИМАНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ПРИЗНАКОВ В НЕЙРОННЫХ СЕТЯХ

В. А. Воробей

Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, v.vorobey.edu@gmail.com
Научный руководитель: А. Э. Малевич, кандидат физико-математических наук, доцент

На основе дискретного вейвлет-преобразования и модели трансформера реализованы несколько сверточных блоков механизма внимания для улучшения извлечения признаков из входных данных при уменьшении размерности. Полученный блок встроен в модель MobileNetV2. Исходная и обновленные модели протестированы на наборе данных LSUN.

Ключевые слова: нейронные сети; дискретное вейвлет-преобразование; механизм внимания; глубокое обучение; компьютерное зрение.

Структура модели

В существующих моделях компьютерного зрения необходимо несколько раз уменьшать размерность входного изображения. Зачастую для этого используются весьма простые слои: субдискретизация по максимальному или среднему значению или же свертка с шагом 2. Предлагается заменить данные слои блоками уменьшения размерности на основе вейвлетов и механизма внимания.

В качестве вейвлетов было решено использовать семейство CDF-9/7 [1], поскольку оно находит широкое применение в области анализа изображений. В частности, используется для сжатия картинок в формате JPEG 2000.

Для механизма внимания выбор был сделан в пользу блоков, которые используются в модели SWin Transformer V2 [2].

Для проведения экспериментов была выбрана модель MobileNetV2 [3], поскольку в настоящее время она является популярным выбором в разработке на мобильных устройствах, показывая высокое быстродействие и уровень качества. В ней 5 блоков уменьшения размерности, причем все для этого используют сверточный слой с шагом 2. Первый блок оставался неизменным, а остальные 4 были заменены. Замена проводилась следующим образом: слой свертки с шагом 2 заменялся на предложенный вариант блока внимания, а после него применялся сверточный слой с шагом 1. Данная схема приведена на рис. 1.

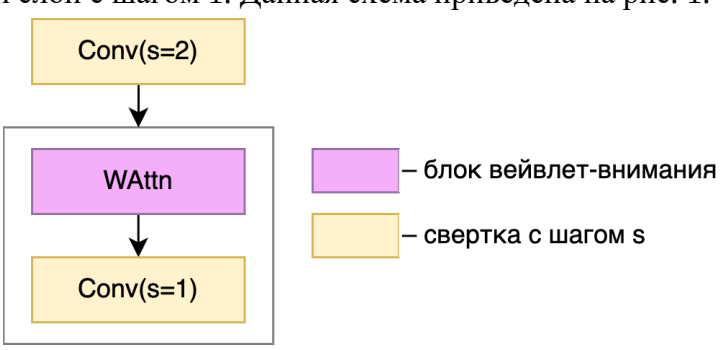

Рис. 1. Схема встраивания блоков WAttn

Было реализовано 2 варианта блока внимания: первый использует низкочастотную компоненту DWT-разложения и блок SWinV2Attn, а второй к выходу первого блока добавляет карту внимания, построенную по высокочастотным компонентам DWT-разложения. Блок

SWinV2Attn использует следующие параметры: глубина – 2, внутренняя размерность скрытого состояние – 72, количество голов – 3, размер окна – 2, размер патча – 4, масштабирующий фактор перцептрона – 4. Для приведения его выхода к нужному размеру применяются слои с транспонированной сверткой с размером ядра 1×1. Также перед блоком SWinV2Attn входной сигнал проходит через несколько уровней DWT-разложения. Таким образом, механизм влияния применяется на карте существенно меньшего разрешения, что значительно увеличивает вычислительную эффективность. Входной сигнал раскладывается до тех пор, пока одна из сторон не будет меньше 16 пикселей. Полученная карта внимания восстанавливается в исходный размер при помощи обратного вейвлет преобразования: на каждом шаге используется обновленная низкочастотная компонента, в то время как высокочастотные компоненты не меняются. Схема блоков представлена на рис. 2.

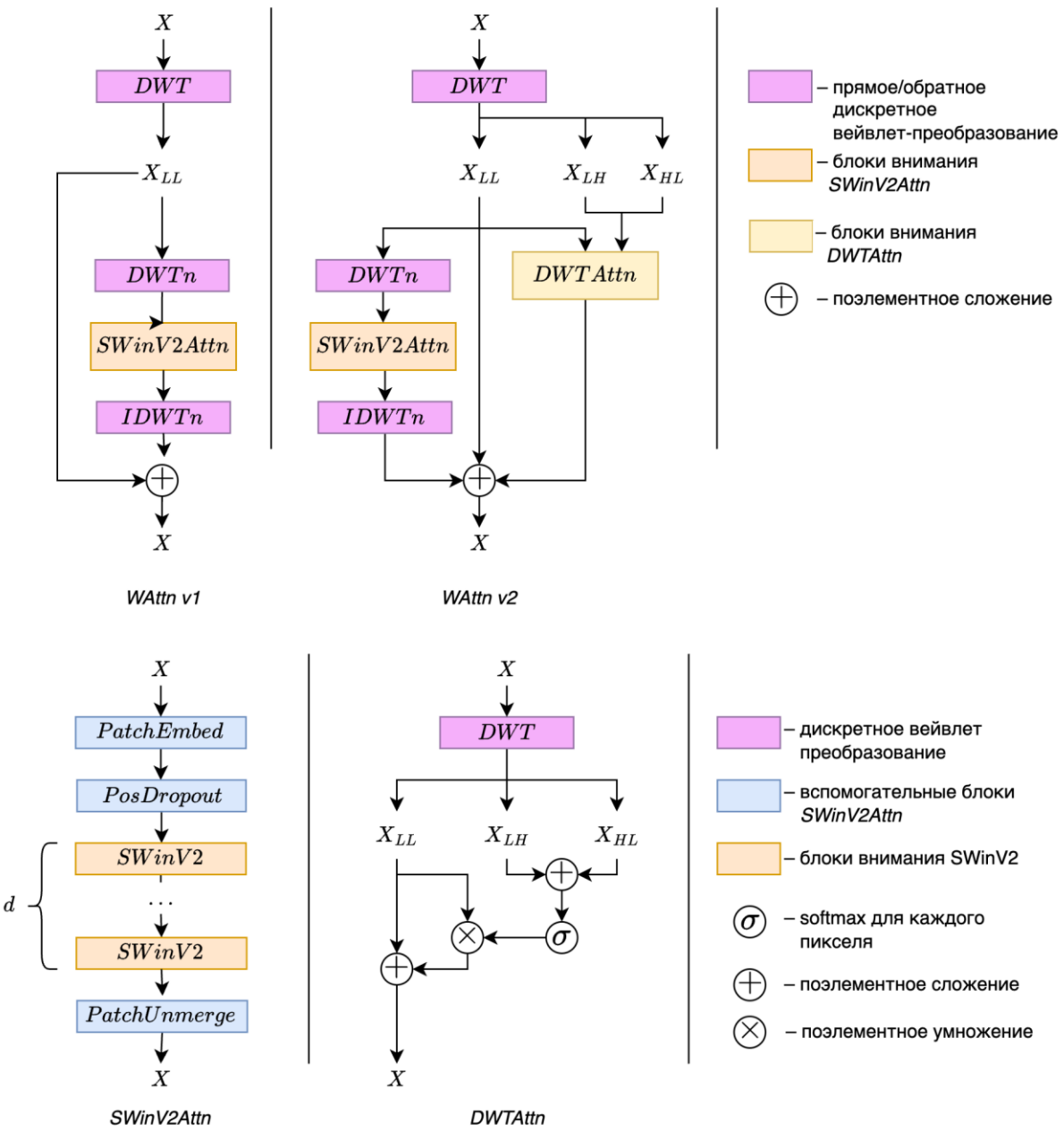

Рис. 2. Схема блоков WAttn v1 и WAttn v2

Описание набора данных

Для тестирования модели использовались данные из набора LSUN. Было выбрано 5 категорий (самолет, корабль, автобус, автомобиль, мотоцикл), для каждой из которых было подготовлено одинаковое количество изображений. Для тренировочных данных брались первые 10 000 изображений из каждой категории, а для тестирования первые 10 000 изображений, начиная со 100 000. Тренировочных данные делились на 2 группы: те, на которых модель непосредственно обучается (85% из каждой категории), и те, на которых проверяется промежуточное качество (15% из каждой категории, валидационный набор). Таким образом, общее количество данных для тренировки, валидации и тестировании составляло 42 500, 7 500 и 50 000 изображений соответственно. Примеры данных для каждой категории представлены на рис. 3.

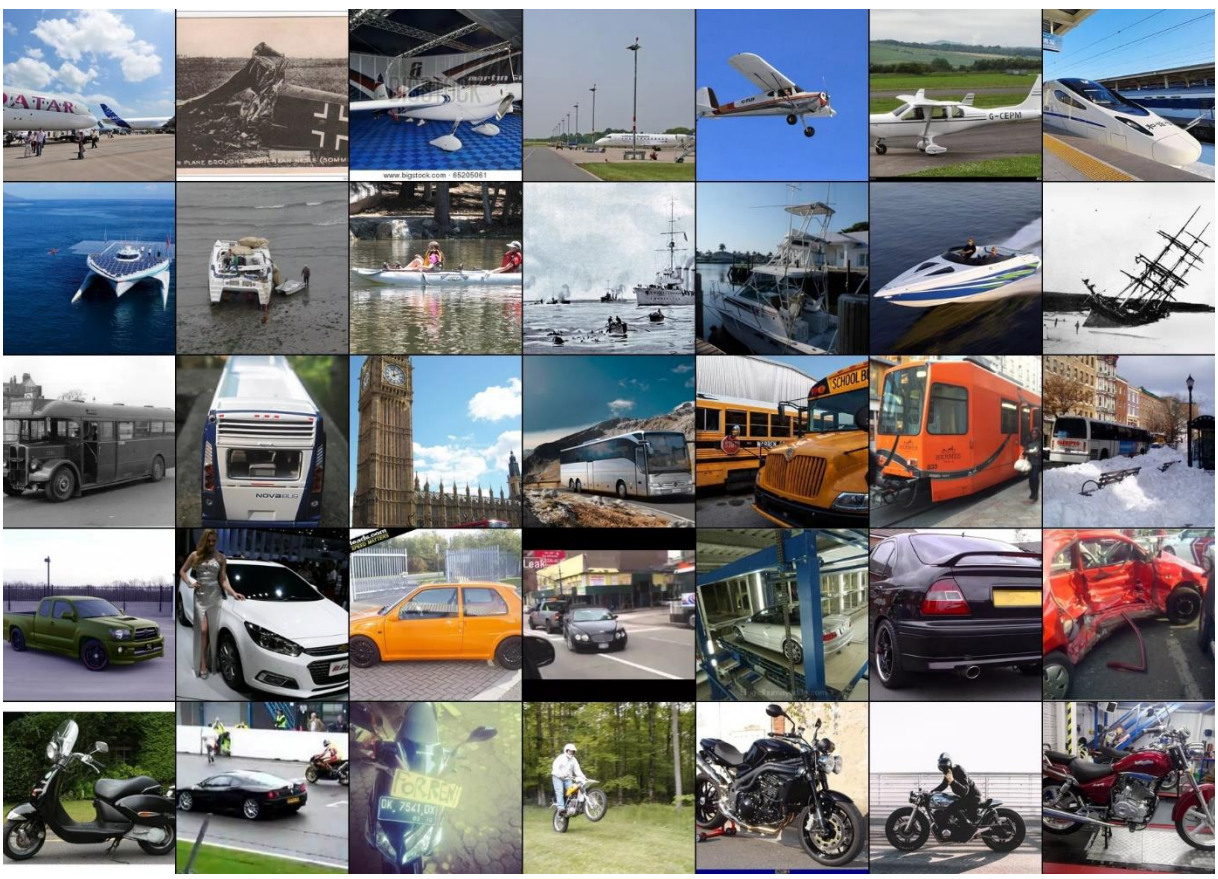

Рис. 3. Примеры изображений из набора данных LSUN

Все изображения заранее приводились к размеру 256x256.

Для тренировочных образцов использовались следующие преобразования: изображения с вероятностью p=0.5 зеркально отражались относительно вертикальной оси, а также с каждой стороны на 15% от ее размера дополнялись нулями, после чего брался случайный патч размером 256×256 пикселей. При проверке качества на валидационном и тестовом наборах данные преобразования отключались.

Результаты

Все модели обучались на протяжении 100 эпох с возможностью ранней остановке, если функция потерь на валидационном наборе не уменьшалась на протяжении 15 эпох.

В следующей таблице приведены результаты для базовой и модифицированных моделей.

Сравнение результатов

Модель	Кол-во параметров (М)	Кол-во операций (ГФЛОПС)	Видеопамять (тренировка, МБ)	Метрика (тест)
MobileNetV2	1,580	0,324	2452	0,9336
MobileNetV2 + WAttn v1	2,316	0,295	2304	0,9206
MobileNetV2 + WAttn v2	2,316	0,295	2368	0,9243

На основании полученных результатов, можно сделать вывод, что существующие варианты блоков уменьшения размерности (например, сверточный слой с шагом 2), показывают весьма хорошую эффективность как по качеству, так и по скорости обучения, превзойти которые не удалось, даже значительно увеличив количество параметров в модели в соответствующих блоках.

Библиографические ссылки

1. Wavelet CDF 9/7 Implementation [Electronic resource] // Getreuer: On Wavelet Implementation, 1997. URL: https://getreuer.info/posts/waveletcdf97/index.html (date of access: 21.10.2023).

2. Swin Transformer V2: Scaling Up Capacity and Resolution / Z. Lie [et al.] // 2021. URL: https://arxiv.org/abs/2111.09883.

3. MobileNetV2: Interested Residuals and Linear Bottlenecks / M. Sandler [et al.] // 2019. URL: https://arxiv.org/abs/1801.04381.

4. LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop / F. Yu [et al] // 2015. URL: https://arxiv.org/abs/1506.03365.