

Модуль самопроизвольного восстановления многопутевого внимания для распознавания дефектов энергетического оборудования на подстанции *

Ифэн Хан, Дунлян Ци, Юньфэн Янь *

Чжэцзянский университет, No.38 ул. Жэда, Ханчжоу 310007 Чжэцзян, Китай

Получено 14 июля 2024 г.; отредактировано 7 октября 2024 г.; принято 7 ноября 2024 г.

Аннотация. Техническое обслуживание энергетического оборудования имеет большое значение в электросетях, где распознавание дефектов на основе обработки изображений призвано классифицировать нештатные состояния во время ежедневного осмотра. Однако из-за размытых признаков на изображениях дефектов существующие алгоритмы распознавания дефектов обладают низкой точностью распознавания. Визуальное внимание может обеспечить точное распознавание благодаря своей способности моделировать долгосрочные зависимости, но при этом вносит дополнительную вычислительную сложность, особенно для многопутевого внимания в визуальных структурах трансформаторов. В этих условиях в данной статье предлагается модуль самопроизвольного восстановления многопутевого внимания, который может снизить вычислительную сложность и легко комбинироваться со сверточной нейронной сетью (CNN). Таким образом, в предлагаемой нами структуре локальные и глобальные признаки могут вычисляться одновременно, что позволяет улучшить производительность распознавания дефектов. В частности, предлагаемый модуль самопроизвольного восстановления многопутевого внимания позволяет уменьшить избыточные параметры, тем самым решая проблему ограниченных вычислительных ресурсов. Экспериментальные результаты были получены на основе набора данных о дефектах, собранных на подстанции. Результаты продемонстрировали эффективность и превосходство предлагаемого метода над другими передовыми алгоритмами. ©2025 Global Energy Interconnection Group Co.Ltd. Издательские услуги предоставлены Elsevier B.V. от имени KeAi Communications Co.Ltd. Данная статья находится в открытом доступе и распространяется по лицензии CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Ключевые слова: многопутевое внимание; распознавание дефектов; энергетическое оборудование; вычислительная сложность

0 Введение

Ежедневный осмотр энергетического оборудования является одним из важнейших методов обеспечения безопасной и стабильной работы энергосистем. Своевременное и быстрое обнаружение дефектов оборудования имеет большое значение для обеспечения безопасности энергосетей. Однако процесс осмотра для тщательного обнаружения дефектов является трудоемким и требует большого объема ручного труда [1]. В последнее время широко изучаются интеллектуальные методы распознавания дефектов на основе обработки изображений для использования в автоматическом осмотре энергетического оборудования. Метод обработки изображений на основе глубокого обучения предназначен для автоматического распознавания существующих дефектов и оповещения дежурного персонала.

В последние годы распознавание изображений достигло значительного прогресса как с помощью CNN, так и с помощью трансформерных структур в области искусственного интеллекта. AlexNet продемонстрировал значительные возможности глубоких сверточных нейронных сетей в задачах классификации изображений [2]. VGGNet сначала оценивает глубокую сеть с использованием сверточных фильтров 3×3, а затем доказывает ее эффективность [3]. ResNet использует остаточную структуру для обеспечения лучших результатов обучения глубоких нейронных сетей [4]. ViT (Vision Transformer) вводит структуру внимания в задачи распознавания изображений и получает самые современные результаты по сравнению со сверточной нейронной сетью [5].

Таким образом, технология распознавания изображений получила широкое применение в интеллектуальных энергосетях. Ли и др. внедрили технологию распознавания лиц в систему доступа интеллектуальной энергосети [6]. Цао и др. исследовали задачи идентификации мелких дефектов на основе технологии распознавания изображений [7]. Чен и др. применили оператор описания признаков SIFT для улучшения алгоритма распознавания изображений интеллектуальной энергосети [8]. В этих условиях улучшилась интеллектуальная проверка энергетического оборудования. Однако, в отличие от общедоступных наборов данных, набор данных о дефектах энергосистемы имеет размытые и сложные характеристики. Это приводит к более жестким требованиям к алгоритмам интеллектуального распознавания изображений.

Существующая интеллектуальная система технического контроля подстанций состоит из трех частей, как показано на рис. 1. Во-первых, ежедневная проверка энергетического оборудования выполняется с помощью инспекционной камеры, которая автоматически обнаруживает дефекты с использованием соответствующего алгоритма обнаружения [9]. После обнаружения дефектов инспекционный робот отправляется в определенное место для детальной повторной проверки, а подробные изображения отправляются обратно на центральный сервер. Затем используется специальный алгоритм распознавания для классификации дефектов. Система подает сигнал тревоги при подтверждении дефекта.

Для повышения точности обнаружения и снижения частоты ложных срабатываний был использован алгоритм распознавания для оценки дефектов на изображениях оборудования, полученных с близкого расстояния. Однако изображения дефектов на подстанции демонстрируют размытые признаки с переменной формой. В этих условиях ожидается, что модель распознавания дефектов будет обладать более мощными возможностями для детального понимания. Кроме того, ежедневные проверки накапливают большое количество изображений, что создает сложные требования к скорости обработки и вычислительной сложности этой модели распознавания.

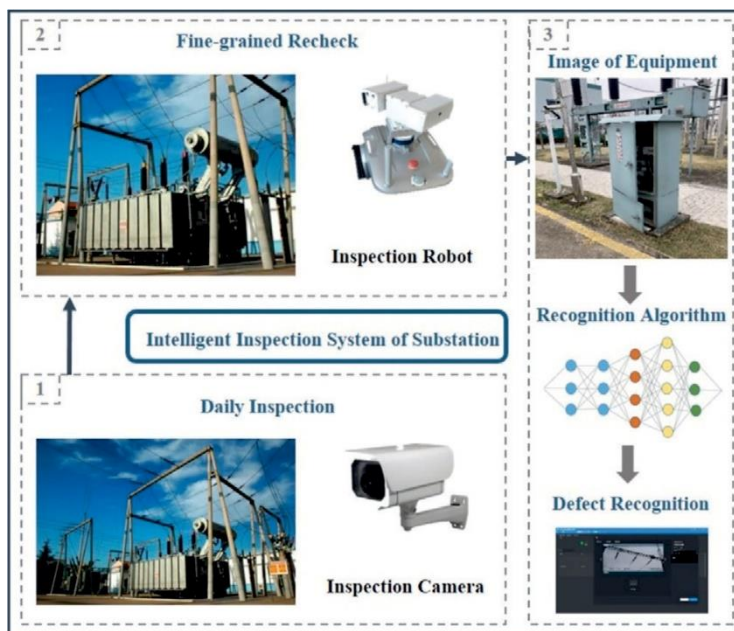


Рис. 1. Интеллектуальная система контроля на подстанции.

Для более детального понимания и повышения точности распознавания исследователи ввели визуальное внимание в изучение задач распознавания изображений. В отличие от структуры CNN, механизм внимания моделирует зависимости на больших расстояниях. Кроме того, модель нейронной сети без механизма внимания обрабатывает все входные признаки изображения одинаково и не может отфильтровывать ключевые признаки. Это затрудняет обучение сети работе с данными изображений с размытыми признаками. Поэтому был разработан механизм

визуального внимания для повышения производительности нейронных сетей с использованием модулей типа «подключи и работай». Сети сжатия и возбуждения (SEnet) [10], модуль внимания сверточного блока (CBAM) [11], нелокальные сети [12] и неявные нелокальные сети [13] повышают точность базовых нейронных сетей на основе различных структур внимания. Впоследствии в задачах распознавания изображений используется визуальный трансформер, в основном состоящий из многопутевых модулей внимания.

Однако задачи распознавания изображений на основе механизма внимания страдают от двух основных проблем. Во-первых, механизм внимания вносит более высокую вычислительную сложность в исходную нейронную сеть. Во-вторых, для конкретного визуального трансформера требуется большое количество высококачественных обучающих данных.

Для получения алгоритма обнаружения дефектов с более высокой способностью детального понимания в данном исследовании создан модуль внимания типа «подключи и работай» и введен механизм самопроизвольного восстановления [13] в многопутевом механизме внимания. Таким образом, был разработан модуль многопутевого внимания с самопроизвольным восстановлением. Этот модуль обладает преимуществом трансформаторной структуры, одновременно решая проблемы более высоких вычислительных затрат и больших требований к данным. Впоследствии модуль многопутевого внимания с самопроизвольным восстановлением был объединен с нейронной сетью для распознавания дефектов энергетического оборудования на подстанциях. Таким образом, этот метод позволяет одновременно моделировать долгосрочную зависимость с помощью многопутевого механизма внимания и глобальные признаки с помощью CNN.

Основные результаты данной работы можно кратко изложить следующим образом:

- 1) Для улучшения детального понимания метода распознавания дефектов был разработан модуль самопроизвольного восстановления многопутевого внимания, который делает многопутевое внимание модулем, готовым к использованию, с меньшим количеством параметров и меньшей вычислительной сложностью.
- 2) Предложенный метод повышает точность задач распознавания дефектов при детальной перепроверке интеллектуальных систем контроля. Ожидается, что это позволит снизить трудозатраты на ежедневную эксплуатацию и техническое обслуживание подстанций.
- 3) Предложенный метод сочетает многопутевое внимание с CNN, объединяющей как локальные, так и глобальные связи. Благодаря специфической конструкции самопроизвольного восстановления многопутевого внимания, этот метод эффективно использует преимущества трансформера и решает проблему сложности обучения, присущую структуре трансформера.

1 Сопутствующие исследования

1.1 Визуальное внимание

В качестве важной части процесса получения внешней информации зрительная система человека обладает точной организационной структурой и сложными механизмами нейронной обработки. Сначала она воспринимает внешние изображения через глаза и преобразует их в соответствующие нейронные сигналы. Столкнувшись с огромным количеством визуальной информации, зрительная система человека может быстро отфильтровать её и точно передать в мозг. Этот тип фильтрации информации, являющийся центральным психологическим регуляторным механизмом человеческого поведения, привлекает внимание исследователей.

В целом, зрительная система человека быстро обрабатывает большой объем внешней визуальной информации с помощью сложного механизма и сохраняет наиболее важные признаки. Вдохновленные этим, исследователи предложили механизмы внимания и применили их в области

глубокого обучения. Механизм визуального внимания позволяет более эффективно извлекать признаки данных и избирательно фильтровать нерелевантную информацию, что имеет важное значение для обучения глубоких нейронных сетей. Механизм визуального внимания моделирует взаимосвязи между входными данными и вводит признаки взаимосвязей в процесс обучения сети, тем самым фокусируя внимание модели на характеристиках данных [14-16].

В задачах обработки изображений исследователи разработали модуль, работающий по принципу «подключи и работай», на основе механизма внимания. SEnet [10] перераспределяет размерности каналов карт признаков, чтобы изменить влияние каждой одноканальной карты признаков на общее распределение признаков. Для достижения интерактивного слияния признаков между канальным и пространственным вниманием, СВМ [11] каскадирует предложенное канальное и пространственное внимание для построения модуля остаточного внимания. Нелокальная сеть [12], типичный представитель моделей пространственного внимания, предназначена для моделирования взаимосвязи внутри каждого канала и присвоения весов внимания различным пикселям. Механизм смешанного внимания высокого порядка [17] предлагает полиномы высокого порядка для прогнозирования сложных взаимосвязей высокого порядка между признаками входного изображения. Многопутевая модель внимания [18] была применена Google к задачам машинного перевода, полностью отказавшись от сверточных нейронных сетей и рекуррентных нейронных сетей с механизмом внимания. Сравнение различных модулей внимания представлено в таблице 1. Среди них многопутевое внимание показало наиболее эффективную производительность. В основе структуры трансформера лежит параллельное соединение нескольких моделей внимания второго порядка, что приводит к ограниченному объему вычислений взаимосвязи и высокой вычислительной сложности.

1.2 Интеллектуальный визуальный осмотр энергетического оборудования на основе механизма визуального внимания

Обнаружение дефектов в энергетическом оборудовании с помощью обработки изображений сталкивается с трудностями, связанными со сложным фоном изображений, малым количеством отрицательных точек данных и нечеткими признаками образцов. В этих условиях исследователи ввели механизмы визуального внимания для интеллектуального контроля задач проверки энергетического оборудования. Чжао и др. [19] добавили механизм внимания к краям в остаточную сеть и предложили сеть генерации на основе внимания для реконструкции низкоразрешенных инфракрасных тепловых изображений энергетического оборудования. Лу и др. [20] предложили метод классификации текста для дефектов энергетического оборудования на основе многопутевой рекуррентной сверточной нейронной сети с вниманием. У и др. [21] использовали модели внимания для достижения слияния признаков во время повышения разрешения, что повысило точность обнаружения ржавчины с помощью энергетического оборудования. Ван [22] объединил сверточный модуль внимания с сетью сегментации и предложил алгоритм распознавания показаний указательного инструмента. Лю и др. [23] предложил алгоритм распознавания табличек электросчетчиков на основе механизма внимания, внедрив механизм внимания в структуру кодировщика-декодера для улучшения результатов распознавания табличек. Чжан [24] объединил алгоритм обнаружения объектов YOLO-v3 с механизмом внимания и предложил алгоритм обнаружения пожара на основе механизма внимания, который повысил выразительность семантики признаков. Хэ и др. [25] объединили сверточные нейронные сети и механизмы внимания каналов для увеличения веса локальной информации о неисправности и достигли повышения точности распознавания при обнаружении молниеотводов, автоматических выключателей, трансформаторов тока и трансформаторов напряжения.

Для достижения более высокой точности автоматической проверки в энергетических системах интеллектуальные алгоритмы должны быть способны к детальному распознаванию. Поэтому большинство алгоритмов используют механизм внимания в этом методе, как показано в таблице 2.

Хотя вышеупомянутые исследователи улучшили возможности интеллектуальной проверки энергетического оборудования, проблемы все еще остаются, такие как недостаточное объединение признаков, высокая вычислительная сложность и трудности в вычислении моделей высокого порядка. Недостаточное объединение признаков указывает на то, что модель просто применяет один тип внимания, вводя в сеть либо канальное, либо пространственное внимание. Кроме того, механизм внимания считается более сложным с точки зрения вычислений из-за вычисления карты сходства. Более того, внимание высокого порядка может моделировать более детальные взаимосвязи, но требует более легкой архитектуры.

Таблица 1. Сравнение нескольких модулей внимания.

Attention module	Relationship function	Order
SNet	Feature in each channel as weight	1
Non-local network	Multiplication between two matrices	2
High-order attention	Multiplication among several matrices	3+
Cascading attention	Spatial attention and channel attention	1 + 2
Multi-head attention	Parallelization of several spatial attention	2

Таблица 2. Сравнение интеллектуальных методов контроля энергетического оборудования.

Method	Advantage	Shortcoming
[19]	Apply edge attention mechanism to the residual network	Insufficient feature fusion
[20]	Combine multi-head attention with recurrent network.	Redundant parameter
[21]	Use attention models to achieve feature fusion during up-sampling	Insufficient feature fusion
[22]	Combine attention module with segmentation network	High computational complexity
[23]	Introduce attention mechanism into the encoder-decoder structure	High computational complexity
[24]	Insert attention module into YOLO-v3 detection method	Difficult for high-order attention calculation
[25]	The application of channel attention	Insufficient feature fusion

В отличие от предыдущих исследований, в данном исследовании сначала создается многопутевой модуль с использованием сверточной функции для моделирования внимания высокого порядка. Затем механизм самопроизвольного восстановления объединяется с предложенным многопутевым механизмом внимания для уменьшения избыточных параметров и вычислительной сложности. Наконец, модуль самопроизвольного восстановления объединяется со сверточной сетью для достижения детального распознавания дефектов на подстанции.

1.3 Трансформатор и многопутевое внимание

В последние годы трансформерные структуры широко используются в обработке естественного языка и компьютерном зрении. В отличие от сверточных нейронных сетей и рекуррентных нейронных сетей, трансформер полностью состоит из механизмов внимания и включает в контекстный вектор характеристики взаимосвязи между семантическими блоками [18].

Чтобы исследовать, может ли механизм внимания полностью заменить сверточные ядра в задачах извлечения признаков изображения, Рамачандран и др. напрямую использовали модуль самовнимания для замены сверточных ядер 3x3 в ResNet и продемонстрировали, что замененная сеть работает лучше [26]. Досовицкий и др. применили трансформер к задачам обработки изображений и предложили модель Vision Transformer [27]. Команда Google предложила

основанную на трансформере структуру обнаружения объектов под названием DETR [28]. Для повышения скорости сходимости и производительности обнаружения объектов модели DETR во время обучения Чжу и др. предложили деформируемый DETR [29]. Модель Pix2seq рассматривает задачи обнаружения объектов в обработке изображений как задачи обработки языка и предполагает, что модель знает категорию и местоположение объекта, обучая модель тому, как выводить информацию во время обучения [30]. В сегментации изображений Segmentation Transformer объединяет ViT с простым декодером для создания мощной модели сегментации [31]. TransUnet встраивает модель трансформера в U-образная сетевая структура [32]. Модель SMT использует как способность модели трансформера извлекать удаленные признаки, так и способность сети извлекать локальные признаки [33]. Для одновременного использования преимуществ сверточных нейронных сетей и механизмов внимания CoAtNet предлагает ряд гибридных моделей механизмов внимания трансформера и сверточной нейронной сети для лучшего баланса количества параметров модели и точности [34]. Модель конформера использует структуру сосуществования трансформера и сверточной нейронной сети для одновременного получения локальных признаков и глобальных зависимостей и достижения слияния двух признаков [35]. Модель CvT вводит сверточные операции в модель ViT, создавая иерархическую структуру трансформера с новыми встраиваниями токенов свертки и блоком сверточного трансформера [36]. Для решения проблемы высокой вычислительной нагрузки в визуальных моделях трансформеров модель P2T применяет пирамидальное пулинг к многопутевым механизмам внимания, одновременно уменьшая длину последовательности и захватывая мощные контекстные признаки [37]. Ли и др. предложили визуальный трансформер с глубоко разделяемым самовниманием для решения проблемы высокой вычислительной нагрузки в моделях трансформеров [38]. Вышеупомянутая модель объединяет основную структуру трансформера с вычислением свертки, используя преимущества обоих алгоритмов, что приводит к созданию более эффективной модели.

Хотя вышеупомянутые модели превосходят существующие алгоритмы в ряде задач обработки изображений, они все же имеют определенные ограничения. Во-первых, сети с многопутевым механизмом внимания в качестве основной структуры страдают от таких проблем, как высокая вычислительная сложность и потребление ресурсов во время обучения. Кроме того, отличная производительность трансформера во многих задачах основана на предварительно обученной модели на больших общедоступных наборах данных, что делает его производительность нестабильной на некоторых редких наборах данных.

2 Метод

В этом разделе сначала предлагается модуль самопроизвольного восстановления многопутевого внимания, а затем он встраивается в ResNet [4] для выполнения задачи распознавания дефектов.

2.1 Модуль самопроизвольного восстановления многопутевого внимания

Как наиболее важный компонент трансформера, многопутевое внимание может моделировать долгосрочные зависимости с помощью нескольких параллельных блоков. В этом разделе для проектирования параллельных блоков используется базовая структура нелокальной сети. Операция самопроизвольного восстановления [13] была расширена до многопутевого внимания.

Структура обучения многопутевого механизма внимания самопроизвольного восстановления показана на рис. 2. Таким образом, в отличие от структуры многопутевого механизма внимания в ViT, каждый блок в предлагаемом многопутевом механизме внимания самопроизвольного восстановления сначала использует операцию свертки для получения векторов признаков. Впоследствии, путем проектирования нескольких параллельных знаменателей, функция вычисления карты сходства может быть автоматически уменьшена после обучения. Входные и выходные карты признаков имеют одинаковый размер, что означает, что модуль

самопроизвольного восстановления МНА может быть легко вставлен на любом этапе глубокой нейронной сети.

Математически входная карта признаков многопутевого модуля внимания с механизмом самопроизвольного восстановления может быть представлена как $X \in \mathbb{R}^{c \times h \times w}$, а выход — как $Y \in \mathbb{R}^{c \times h \times w}$, где c — канал; h и w — ширина и высота карты признаков соответственно; и n — номер головки многопутевого модуля внимания, который указывает количество отдельных блоков, полностью параллельных друг другу.

Сначала входной сигнал X проходит через сверточный слой f_{conv} с ядром 1×1 и получает карту признаков

$$X' \in \mathbb{R}^{\frac{c}{2} \times h \times w}:$$

$$X' = f_{\text{conv}}(X) \quad (1)$$

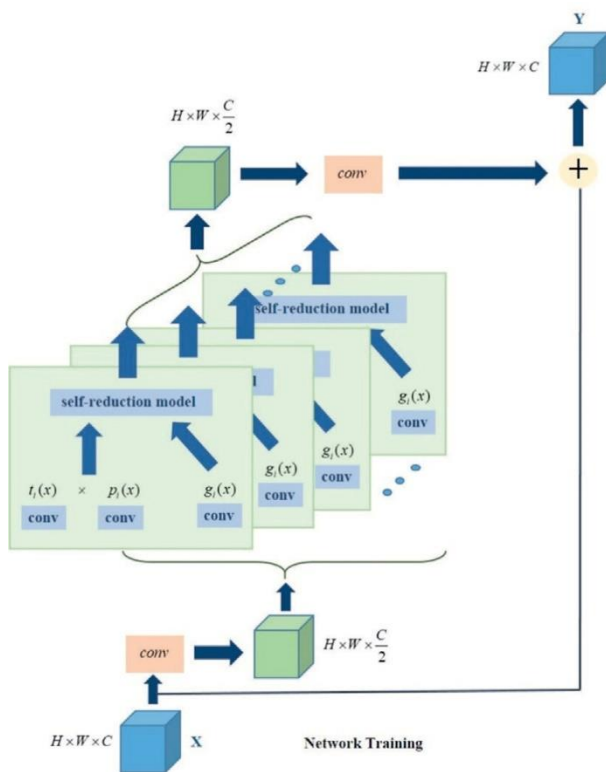


Рис. 2. Структура самопроизвольного восстановления многопутевого внимания.

Затем X' подается на вход n многопутевых модулей внимания, в каждом модуле:

X' пропускается через три сверточных слоя с ядром 1×1 для получения трех выходных признаков t_i

(x), p_i (x) и g_i (x) размером $\frac{c}{2n} \times h \times w$, и $i \in [1, n]$. Затем эти три характеристики

$$1 \times \frac{hwc}{2n}$$

расширяются до трех двумерных карт характеристик размером $1 \times \frac{hwc}{2n}$ и yield θ_1 (x), θ_2 (x), и g (x). Аналогично процессу вычислений в [13], предлагаемый метод сначала устанавливает знаменатель самопроизвольного восстановления для каждого блока. Градиенты

этих знаменателей не вычисляются в процессе обучения обратного распространения ошибки. Во время вывода эти знаменатели служат константами.

Математический процесс можно проиллюстрировать следующим образом:

$$M_i = \frac{\sum_{i,j} (\theta_1(x_{i,j}) \odot \theta_2(x_{i,j})) \times g(x)}{\text{value}[\sum_{i,j} (\theta_1(x_{i,j}) \odot \theta_2(x_{i,j}))]} \quad (2)$$

M_i представляет собой результат работы каждого простого блока. значение $[\sum_{i,j} (\theta_1(x_{i,j}) \odot \theta_2(x_{i,j}))]$ представляет собой знаменатель, который должен обеспечить самопроизвольное восстановление в процессе вывода.

После этого эти выходные тензоры объединяются, и получается карта

$$Y' \in R_2^{c \times h \times w}$$

признаков :

$$Y' = f_{\text{concat}}(M_i) \quad (3)$$

Затем Y' проходит через сверточный слой с ядром 1×1 и добавляется к входному X для построения следующей формы остатка:

После обучения параметры $t_i(x)$ и $p_i(x)$ могут быть автоматически уменьшены. Структура показана на рис. 3. Процедура вывода может быть описана математически:

Сначала,

$$X'_{\text{inference}} = f_{\text{conv}}(X) \quad (5)$$

$$M_i = g_i(X'_{\text{inference}}) \quad (6)$$

Затем полученные результаты объединяются:

$$Y'_{\text{inference}} = f_{\text{concat}}(M_i) \quad (7)$$

Итоговый результат:

$$Y = f_{\text{conv}}(Y'_{\text{inference}}) + X \quad (8)$$

2.2 Метод интеллектуального распознавания дефектов

Современные интеллектуальные системы контроля на подстанциях полагаются на высококачественные алгоритмы с детальным пониманием и способностью извлечения признаков. В данном исследовании был разработан модуль самопроизвольного восстановления МНА для повышения производительности модуля распознавания дефектов. Многопутевое внимание является ключевым компонентом трансформатора и считается наиболее мощной глубокой нейронной сетью в ряде задач компьютерного зрения. Однако обучение трансформатора требует большого объема высококачественных данных. Это затрудняет его применение в задачах интеллектуального контроля для интеллектуальных энергосетей. В отличие от этого, сверточная нейронная сеть не может моделировать долгосрочные зависимости, как трансформатор, но требует меньше обучающих данных. В этих условиях в данном исследовании была проведена комбинация МНА и сверточной нейронной сети.

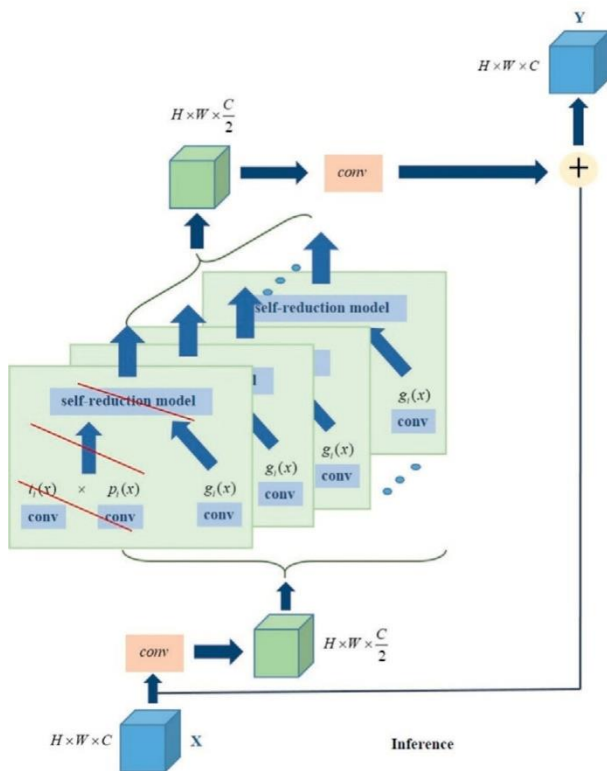


Рис. 3. Структура самопроизвольного восстановления многопутевого внимания после тренировки.

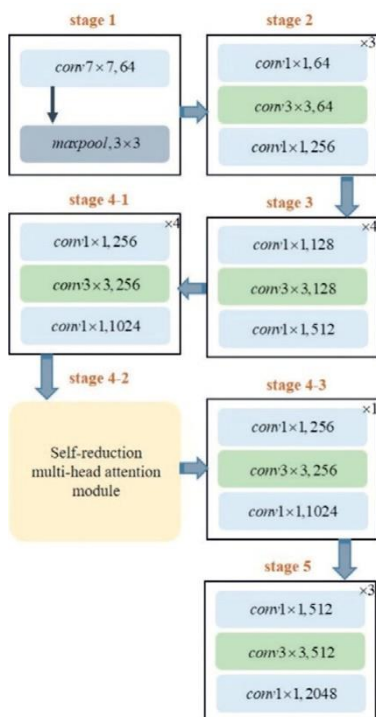


Рис. 4. ResNet50 с предложенным механизмом многопутевого внимания.

В частности, этот метод вставляет предложенный модуль многопутевого внимания с функциями самопроизвольного восстановления в ResNet50[4]. Структура интеллектуального метода распознавания дефектов показана на рис. 4 и 5. Используя предложенный многопутевой механизм внимания, сеть может моделировать долгосрочные зависимости и вводить дополнительную информацию о взаимосвязях в процессе обучения. Таким образом, модель может извлекать более подробную информацию из обучающего набора данных.

3 Испытания

3.1 Обучающий набор данных

Для оценки эффективности предложенного метода мы создали набор данных для распознавания дефектов. Все изображения были собраны на подстанции электросети, что обеспечило подлинность и достоверность обучающих данных. Обучающий набор данных включает четыре категории: некорректное закрытие электрощитка, размытая крышка счетчика, изменение цвета силикагеля и разрыв изолятора. Примеры этих обучающих изображений показаны на рис. 6.

Распределение наборов данных о дефектах приведено в таблице 3. Набор данных состоит из 3772 изображений, включая 3192 обучающих и 580 тестовых изображений. Все изображения представляют собой крупные планы деталей дефекта. Размер каждого изображения был изменен до 608×608 пикселей.

3.2 Испытания в области вычислительной сложности

Для оценки способности нашего метода снижать вычислительную сложность при одновременном повышении точности распознавания дефектов, сначала в этом разделе оцениваются операции с плавающей запятой (FLOPs) и параметры между исходным многопутевым механизмом внимания и предложенным нами модулем. FLOPs использовались для измерения общей сложности (требуемой вычислительной нагрузки). В этом испытании количество головок многопутевого механизма внимания было установлено равным восьми.

В данном исследовании входные карты признаков для многопутевого модуля внимания и предложенного нами модуля были установлены на значения 64×64 и 128×128 соответственно. Результаты эксперимента представлены в таблице 4, где 1GFLOPs = 1×10⁹ FLOPs. Результаты оценки доказывают, что предложенный нами метод имеет меньшую вычислительную сложность и меньшее количество параметров, чем исходный многопутевой модуль внимания.

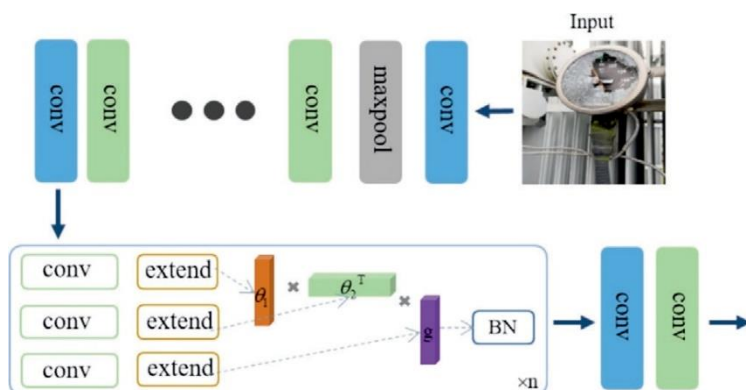


Рис. 5. Структурная схема предлагаемого метода распознавания.

3.3 Испытания по точности распознавания дефектов

Для проверки эффективности предложенного метода многопутевого внимания и распознавания дефектов сеть была обучена с использованием предложенного набора данных о дефектах. Как показано на рис. 4, многопутевое внимание самопроизвольного восстановления объединено с ResNet50 на этапе 4-2. Для оценки эффективности предложенного метода в этом исследовании предложенный метод заменен на SEnet, нелокальную сеть, CBAM и многопутевое внимание из структуры трансформера. SEnet применяет к сети внимание к каналам, тогда как нелокальные сети используют пространственное внимание. CBAM последовательно соединяет пространственное и каналное внимание. Кроме того, предложенный метод был сравнен с двумя другими методами, которые фокусируются на оптимизации механизма самовнимания. Сеть перекрестного внимания

(CCnet) [39] применяет перекрестное внимание для снижения вычислительной сложности, а сеть глобального контекста (GCnet) [40] упрощает нелокальную сеть за счет упрощенного вычисления карты сходства.

Обучение проводилось с использованием видеокарты GeForce RTX 3090. Размер пакета данных был установлен на 64, а скорость обучения — на 1×10^{-3} . Процесс обучения длился 100 эпох (периодов). В качестве основной метрики оценки в задачах распознавания изображений TOP1 представляет собой точность результатов распознавания, когда сеть выдает результат с наибольшей вероятностью в качестве окончательного ответа.

Согласно экспериментальным результатам, представленным в таблице 5, предложенный метод распознавания дефектов превзошел другие алгоритмы с различными модулями внимания. Это объясняется тем, что предложенный метод может одновременно моделировать долгосрочные зависимости с помощью многопутевого механизма внимания и глобальные признаки из сверточной нейронной сети (CNN). Для сравнения, CCnet и GCnet удаляют несущественные ветви во время обучения, что приводит к потере информации. Кроме того, по сравнению с другими упрощенными методами на основе трансформеров, наш предложенный метод фокусируется на разработке подключаемого многопутевого модуля внимания вместо прямого использования структуры трансформера. Таким образом, наш метод позволяет улучшить детальное понимание исходной модели с меньшей вычислительной сложностью. Кроме того, наш предложенный метод использует меньшее количество обучающих данных, что делает его более эффективным, чем методы на основе трансформеров.



Рис. 6. Примеры изображений дефектов в предлагаемом наборе данных.

Таблица 3. Распределение набора данных о дефектах на подстанции.

Defect in substation	Train	Test
Blurred meter cover	884	160
Silica gel discoloration	924	140
Insulator rupture	700	140
Abnormal closure of electricity box	684	140
Total	3,192	580

Таблица 4. Сравнение вычислительной сложности.

Input	Model	Self reduction	FLOPs	Parameters
64×64	Multi-head attention	×	0.47G	72,192
64×64	Ours	✓	0.34G	68,096
128×128	Multi-head attention	×	1.89G	72,192
128×128	Ours	✓	1.35G	68,096

3.4 Исследование абляции

Для проверки воспроизводимости предложенного нами метода на других наборах данных мы оценили различные модели для наборов данных CUB [41] и Cifar100 [42]. Набор данных CUB состоит из 200 типов птиц с 11 788 изображениями, в котором обучающий набор содержит 5994 изображения, а тестовый набор — 5794 изображения. Cifar100 состоит из 100 типов данных, каждый из которых содержит 600 изображений. В этом наборе данных каждый тип данных содержал 500 обучающих и 100 тестовых изображений. Обучение проводилось с использованием GeForce RTX 3090. Размер пакета был установлен на 64, а скорость обучения — на 1×10^{-3} . Процесс обучения длился 110 эпох (периодов).

В таблице 6 представлены результаты оценки различных модулей внимания с использованием набора данных CUB. В таблице 7 представлены результаты тестирования с использованием набора данных Cifar100.

Таблица 5. Результаты обнаружения дефектов предлагаемого метода.

Model	Attention	TOP1
ResNet50	×	81.38
ResNet50 + SEnet	Channel attention	84.50
ResNet50 + non-local network	Spatial attention	85.58
ResNet50 + CBAM	Cascade attention	82.76
ResNet50 + multi-head attention	Multi-head attention	88.26
GCnet	Cascade attention	86.12
CCnet	Criss-cross attention	85.42
Ours	Self-reduction multi-head attention	90.12

Таблица 6. Результаты тестирования модулей внимания с использованием набора данных CUB.

Model	Attention	TOP1
ResNet50	×	85.87
ResNet50 + SEnet	Channel attention	85.95
ResNet50 + non-local network	Spatial attention	85.97
ResNet50 + multi-head attention	Multi-head attention	86.35
Ours	Self-reduction multi-head attention	86.75

Таблица 7. Результаты тестирования модулей внимания с использованием набора данных cifar100 (без предварительно обученной модели).

Model	Attention	TOP1
ResNet50	×	73.24
ResNet50 + SEnet	Channel attention	73.41
ResNet50 + non-local network	Spatial attention	73.53
ResNet50 + multi-head attention	Multi-head attention	74.27
Ours	Self-reduction multi-head attention	74.87

Результаты исследований демонстрируют эффективность и воспроизводимость предложенного модуля самопроизвольного восстановления многопутевого внимания с использованием других

наборов данных, за исключением предложенного набора данных для распознавания дефектов в электросетях.

3.5 Анализ ограничений

В этом разделе анализируются ограничения нашего исследования на основе уровней данных и модальности:

Уровень данных: в данном исследовании основное внимание уделялось детальному распознаванию дефектов во время ежедневных проверок на подстанции. Однако обучающий набор данных был собран в тот же день. Инфракрасные изображения, полученные ночью, были исключены из обучающего набора данных. В будущем наш метод может быть расширен для одновременного применения как к видимым (дневным), так и к инфракрасным (ночным) изображениям.

Уровень модальности: многомодальная обработка данных позволяет достичь информационной взаимодополняемости, что полезно для детального распознавания дефектов. К таким модальностям относятся изображения, физические сигналы (ток и напряжение) и исторические текстовые записи. Предложенный метод не учитывает методы обработки многомодальных входных данных, что может быть улучшено в будущем.

4 Заключение

Точное распознавание имеет большое значение для задач интеллектуального контроля в интеллектуальных энергосетях. Своевременные предупреждения об интеллектуальном распознавании дефектов могут помочь обеспечить безопасную и стабильную работу энергосистемы. Однако существующие алгоритмы сталкиваются с рядом проблем. Во-первых, существующие методы, основанные на механизме внимания, используют либо канальное, либо пространственное внимание, что приводит к недостаточному объединению признаков. Во-вторых, применение механизма внимания вводит избыточные параметры и требует больших вычислительных ресурсов, что затрудняет его применение на периферийных устройствах. Кроме того, существующие алгоритмы не могут моделировать внимание высокого порядка из-за отсутствия упрощенной структуры внимания.

В этих условиях в данном исследовании впервые предлагается модуль самопроизвольного восстановления многопутевого внимания, который позволяет снизить вычислительную сложность и одновременно повысить точность распознавания. В частности, каждый компонент модуля обеспечивает глобальное слияние признаков из каждой пиксельной точки. Многопутевое внимание позволяет легко моделировать зависимости высокого порядка по сравнению с канальным или пространственным вниманием. Предложенный модуль был объединен с классической сверточной нейронной сетью ResNet50 и обучен для достижения детального распознавания дефектов на подстанции. Конструкция самопроизвольного восстановления уменьшает избыточные параметры и снижает вычислительную сложность. Экспериментальные результаты показывают, что предложенный метод имеет меньшую сложность и превосходит другие передовые методы.

Ежедневная проверка на подстанции выявляет аномальные участки с помощью стационарных камер, а затем определяет конкретную категорию дефекта на основе алгоритмов детального распознавания. Возможности алгоритмов детального распознавания определяют уровень их интеллекта в процессе автоматической проверки. Кроме того, алгоритмы детального распознавания дефектов для энергетического оборудования обычно устанавливаются на дронах или роботах-контролёрах, что требует низкой вычислительной сложности и ограниченных вычислительных ресурсов. Разработанный в данном исследовании алгоритм повышает

способность обеспечивать детальное понимание при одновременном снижении вычислительной сложности. Это позволяет модели более эффективно осуществлять детальную классификацию дефектов и имеет практическое значение для построения интеллектуальных энергосетей.

Для повышения точности распознавания алгоритмом в интеллектуальных энергосетях в данном исследовании метод самопроизвольного восстановления сочетается с механизмом многопутевого внимания. Однако этот метод фокусируется только на одной модальности в течение дня, не учитывая многомодальное слияние. Поскольку многомодальная информация более полезна для точной классификации, предложенный модуль внимания может быть дополнительно исследован для обработки кросс-модальной информации. Наш метод может быть дополнительно улучшен для одновременной обработки изображений в видимом и инфракрасном диапазонах.

О вкладе авторов в проект CRediT

Ифэн Хань: написание оригинального текста, методология, исследование, концептуализация. **Дунлянь Ци:** проверка достоверности, научное руководство. **Юньфэн Янь:** написание обзора и редактирование, формальный анализ.

О конфликте интересов

Авторы заявляют, что у них нет известных конкурирующих финансовых интересов или личных отношений, которые могли бы повлиять на результаты работы, представленной в данной статье.

Благодарности

Эта работа частично поддержана в рамках основной программы Национального фонда естественных наук Китая (грант 62127803).

Список используемой литературы

- [1] Р.Х. Цзяо, Ю.З. Лю, Х.Х. и др., Модель глубокого обучения для обнаружения дефектных компонентов малого размера в опорах линий электропередачи, *IEEE Trans.Power Delivery* 37 (4) (2022) 2551-2561.
- [2] А. Крижевский, И. Суцкевер, Г.Е. Хинтон, Классификация ImageNet с помощью глубоких сверточных нейронных сетей, *Commun.ACM* 60 (6) (2017) 84-90.
- [3] К. Симонян, А. Зиссерман, Очень глубокие сверточные сети для крупномасштабного распознавания изображений, *arXiv: 1409.1556*, 2014.
- [4] К.М. Хе, С.Ю. Чжан, С.К. Рен и др., 2016. Глубокое остаточное обучение для распознавания изображений, в: *Труды конференции IEEE по компьютерному зрению и распознаванию образов (CVPR)*, 2016, Лас-Вегас, США, стр.: 770-778.
- [5] А. Досовицкий, Л. Бейер, А. Колесников и др., Изображение стоит 16x16 слов: Трансформеры для распознавания изображений в масштабе. Предварительная публикация *arXiv: 2010.11929*, 2020.
- [6] З.П.Ли, Дж.Ван, С.Шен и др., Проектирование системы доступа к сети интеллектуальной энергосистемы на основе распознавания лиц, в: *Труды 4-й Международной конференции IEEE по силовой электронике и компьютерным приложениям (ICPECA)*, 2024, Шэньян, Китай, 2024, стр. 1238-1242.
- [7] Г. Цао, Ю. Лю, З. Фан и др., Исследование идентификации и обнаружения мелких дефектов в линиях электропередачи интеллектуальной энергосистемы на основе распознавания изображений,

- в: Труды 4-й Международной конференции IEEE по автоматизации, электронике и электротехнике (AUTEEE), 2021, Шэньян, Китай, 2021, стр. 423-427.
- [8] Л. Чен, В. И, Л. Х. Чжан и др., Распознавание изображений в интеллектуальной энергосети на основе нейронной сети и алгоритма SIFT, в: Труды Международной конференции по сетям, информатике и вычислениям (ICNETIC), 2023, Палермо, Италия, 2023, стр. 1-5.
- [9] Ч. Хуан, М. Х. Чен, Л. Ван, Полуконтролируемое обнаружение дефектов поверхности лопастей ветротурбин с помощью YOLOv4, *Global Energy Interconnect*.7 (3) (2024) 284-292.
- [10] Дж. Ху, Л. Шен, Г. Сунь, Сети сжатия и возбуждения, в: Труды конференции IEEE/CVF по компьютерному зрению и распознаванию образов, 2018, Солт-Лейк-Сити, 2018, стр. 7132-7141.
- [11] С. Ву, Дж. Парк, Дж. Й. Ли и др., СВММ: модуль внимания сверточных блоков, в: Сборник лекций по информатике, Издательство Springer International Publishing, 2018, стр. 3-19.
- [12] Х. Л. Ван, Р. Гиршик, А. Гупта и др., Нелокальные нейронные сети, в: Труды конференции IEEE/CVF по компьютерному зрению и распознаванию образов. Солт-Лейк-Сити, 2018, Юта, США, 2018, стр. 7794-7803.
- [13] Ю. Ф. Хан, С. Чен, С. Дж. Чжан и др., iNL: неявная нелокальная сеть, нейрокомпьютерные системы 482 (2022) 50-59.
- [14] Р.Б.Ли, К.Сянь, Ч.Х.Шен и др., Глубокая классификационная сеть на основе внимания для надежного прогнозирования глубины, в: Сборник лекций по информатике, Издательство Springer International Publishing, 2019, стр. 663-678.
- [15] Н.Сарафианос, С.Сюй, И.А.Какадиарис, Глубокая несбалансированная классификация атрибутов с использованием агрегации визуального внимания, в: Сборник лекций по информатике, Издательство Springer International Publishing, 2018, стр. 708-725.
- [16] Х.В.Ге, З.Х.Ян, В.Х.Ю и др., Сверточная сеть LSTM на основе механизма внимания для распознавания действий в видео, *Мультимедийные инструменты Appl*. 78 (14) (2019) 20533-20556.
- [17] Б.Х.Чен, В.Х.Денг, Дж.Н.Ху, Смешанная сеть внимания высокого порядка для повторной идентификации человека, в: Труды Международной конференции IEEE/CVF по компьютерному зрению (ICCV), 2019, Сеул, Республика Корея, 2019, стр. 371-381.
- [18] А.Васвани, Н.Шазир, Н.Пармарх и др., Внимание — это все, что вам нужно, *Adv.NeuralInf.Process.Syst.* (2017).
- [19] Х.С. Чжао, Ю.Х. Пэн, Б.К. Лю и др., (2022) Сверхразрешающая реконструкция тепловизионных изображений электрооборудования на основе генеративно-состязательной сети с вниманием к краям, *Proc.CSEE* 42(10) (2022) 3564-3573.
- [20] С.Х.Лу, Классификационная модель текстов записей о дефектах энергетического оборудования на основе многопутевого внимания сети RCNN, Университет Гуанси, 2021.
- [21] З.Х.Ву, В.Х.Сюн, Дж.Ф.Рен и др., Обнаружение объектов коррозии на силовом оборудовании на основе легковесного SSD, *Comput.Syst.Appl.* 29 (2) (2020) 262-267.
- [22] Дж.Л.Ван, Исследование и применение алгоритма распознавания изображений силового оборудования на основе глубокого обучения, Чжэцзянский университет, 2021.
- [23] Ю. Лю, З.Б. Чжан, В. Чжан и др., Идентификация табличек с характеристиками силовых устройств на основе модели визуального внимания, *Китайский журнал «Электронные устройства»* 45 (3) (2022) 623-627.

- [24] Х.Б. Чжан, Разработка программного обеспечения для распознавания неисправностей на основе многоисточниковых изображений для мониторинга электрооборудования и окружающей среды, Юго-восточный университет, 2021.
- [25] К.Ю.Хе, Д.Т.Пан, Г.Л.Ли и др., Система мониторинга энергетического оборудования на основе модели распознавания изображений SE-Attention, Северо-западный политехнический университет 35 (4) (2021) 71-76.
- [26] П. Рамачандран, Н. Пармар, А. Васвани и др., Автономное самовнимание в моделях зрения, в: NeurIPS, 2019, стр. 68-80.
- [27] А. Досовицкий, Л. Бейер, А. Колесников и др., Изображение стоит 16x16 слов: Трансформеры для распознавания изображений в масштабе, в: ICLR, 2020.
- [28] Х.Г. Сурияге, Х. Ратнайаке, Сквозное обнаружение объектов с помощью трансформеров: Дополнительные материалы, в: ECCV, 2020, стр. 213-229.
- [29] Х.З.Чжу, В.Ж.Су, Л.В.Лю и др., Деформируемые трансформеры для сквозного обнаружения объектов, в:ICLR, 2020.
- [30]Т.Чень, С.Саксена, Л.Ли и др., Pix2seq: Структура языкового моделирования для обнаружения объектов.arXiv: 2109.10852, 2021.
- [31] С.Х. Чжэн, Дж.К. Лу, Х.С. Чжао и др., Переосмысление семантической сегментации с точки зрения перехода от последовательности к последовательности с помощью трансформатора, в: Труды конференции IEEE/CVF по компьютерному зрению и распознаванию образов (CVPR), 2021, Нашвилл, США, 2021, стр.: 6881-6890.
- [32] Дж.Н. Чен, Ю.Ю. Лу, К.Х.Ю и др., TransUNet: трансформеры создают надежные кодировщики для сегментации медицинских изображений: 2102.04306, 2021.
- [33] Дж.Ю.Гуо, К.Хан, Х.Ву и др., СМТ: сверточные нейронные сети и преобразователи зрения, в: Труды конференции IEEE/CVF по компьютерному зрению и распознаванию образов (CVPR), 2022, Новый Орлеан, США, 2022, стр. 12175-12185.
- [34] З.Даи, Х.Лиу, К.В.Ле и др., Coatnet: Сочетание свертки и внимания для всех размеров данных, Adv.NeuralInf.Proces.Syst.34(2021)3965-3977.
- [35]З.Л.Пенг, В.Хуанг, С.З.Гу и др., Conformer: Локальные признаки, связывающие глобальные представления для визуального распознавания, в: Международная конференция IEEE/CVF по компьютерному зрению (ICCV), 2021, Монреаль, Канада, 2021, стр. 367-376.
- [36] Х.П.Ву, Б.Сиао, Н.Кодэлла и др., CvT: введение сверток в графические трансформеры, в: Труды Международной конференции IEEE/CVF по компьютерному зрению (ICCV), 2021, Монреаль, Канада, 2021, стр. 22-31.
- [37] Ю.Х.Ву, Ю.Лту, Х.Жан и др., P2T: пирамидальный пулинговый трансформер для понимания сцены, IEEE Trans.Pattern Anal.Mach.Intell.45 (11) (2023) 12760-12771.
- [38]В.Ли, Х.Ванг, Х.Сиа и др., SepViT: Разделяемый трансформер зрения: 2203.15380, 2022.
- [39]З.Л.Хуанг, Х.Ж.Ванг, Л.С.Хуанг и др., CCNet: Перекрестное внимание для семантической сегментации, в: Труды Международной конференции IEEE/CVF по компьютерному зрению (ICCV), 2019, Сеул, Республика Корея, 2019, стр. 603-612.

[40] Ю. Сао, Ж. Р. Сю, С. Лин и др., GCNet: Нелокальные сети встречаются с сетями сжатия-возбуждения и не только, в: Труды Международной конференции IEEE/CVF по компьютерному зрению (ICCVW), 2019, Сеул, Республика Корея, 2019.

[41] С. Ва, С. Брансон, П. Велиндер и др., Набор данных caltech-ucsd птиц -200-2011, 2011.

[42] А. Крижевский, Г. Хинтон, Изучение нескольких слоев признаков из крошечных изображений, 2009.



Отсканируйте для получения более подробной информации

Рецензирование осуществляется компанией Global Energy Interconnection Group Co.Ltd.

* Ответственный автор

Адреса электронной почты: hanyf@zju.edu.cn (Ю.Хан), qidl@zju.edu.cn(Д.Ци), 21210004@zju.edu.cn (Ю.Янь).

Эта статья переведена из журнала <Global Energy Interconnection> (ISSN: 2096-5117), выпуск 1, 2025 г. Оригинальное название статьи:< Self-reduction multi-head attention module for defect recognition of power equipment in substation>. Перевод предоставляется исключительно для справки; преимущественную силу имеет оригинал: <https://doi.org/10.1016/j.gloi.2024.11.016>.

2096-5117/© 2025 Global Energy Interconnection Group Co.Ltd. Издательские услуги предоставлены Elsevier B.V. от имени KeAi Communications Co.Ltd.

Данная статья находится в открытом доступе и распространяется по лицензии CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Ифэн Хань получил степень бакалавра технических наук в области автоматизации в 2017 году в Чжэцзянском университете. В 2018 году он получил степень магистра технических наук в области систем управления в Имперском колледже Лондона, а в 2024 году — степень доктора философии в Чжэцзянском университете. В настоящее время его научные интересы охватывают обработку изображений и искусственные нейронные сети.



Дунлянь Ци получила степень доктора философии в Школе электротехники Чжэцзянского университета (Китай) в 2002 году. В настоящее время она является профессором и научным руководителем аспирантов в Чжэцзянском университете. Ее последние научные интересы охватывают интеллектуальную обработку информации, хаотические системы, а также нелинейную теорию и ее применение.



Юньфэн Янь получила степень доктора философии в Институте электротехники Чжэцзянского университета (Китай) в 2019 году. В настоящее время она является доцентом Чжэцзянского университета. Ее последние научные интересы охватывают интеллектуальные энергетические системы и технологии обработки изображений.

(Редактор: Цзэдун Чжан)