Название статьи (авторы статьи):
Are Large-scale Datasets Necessary for Self-Supervised Pre-training? (12/20/2021) by Alaaeldin El-Nouby, et al.

Автор рецензии: Хамдеева Дилара, 181

1. Содержание и вклад: опишите статью и вклад авторов в двух-трех предложениях.
    a. Статья отвечает на вопрос: нужны ли огромные датасеты типа ImageNet для self-supervised pretraining?
    b. Ответ на который: нет, если мы используем denoising autoencoders.
    '*We believe that large scale datasets, such as ImageNet, are not necessary for self-supervised pre-training when using denoising autoencoders.'*
    c. Вклад. Авторы утверждают, что предложенный метод влечет за собой следующие важные свойства:
        i. "denoising autoencoders are more **sample efficient**" – можно хорошо претрейниться на меньших объемах данных
        ii. возможность "***to pre-train directly on the target task data*** and obtain a competitive performance". – что позволяет избежать domain shift, который возникает когда мы претрейнимся на X а файнтюнимся на У.
        iii. возможность достичь хороших результатов даже на "***non object-centric images such as COCO"***.
2. Сильные стороны: опишите сильные стороны статьи. Критерии оценки, как правило, включают: корректность утверждений (теоретическая обоснованность, полнота эмпирического анализа), значимость и новизна вклада, актуальность для исследовательского сообщества.
    a. корректность утверждений
        i. Тк. я не очень разбираюсь в этой области мне было трудно оценить корректность утверждений. Но на мой взгляд, работа имеет довольно полный эмпирический анализ:
            1. присутствуют результаты и оценка работы метода (SplitMask) на задачах – image classification, object detection, segmentation
            2. помимо экспериментов поставленных на каких-то определенных задачах (классификация, детекция, сегментация) также приводится анализ, почему метод, который предлагают авторы, хорош:
                a. исследование влияния размеров сэмплов
                b. сравнение denoising encoders с joint embeddings
                c. исследование результатов работы на non-centric images
                d. сравнение предложенного метода (SplitMask) с существующими (BEiT)
            3. наличие подробных графики с полным описанием, как в кэпшене так и в тексте статьи + еще они понятные))
    b. значимость и новизна вклада
        i. значимость и новизна:
            1. утверждается, что статья отвечает на ранее открытые вопросы в области применения трансформеров для картинок, а именно:
                a. дает возможность предобучаться на более вариативной выборке:
                    i. датасеты меньших размеров, чем Imagenet;
                    ii. non object-centric images
                b. а также позволяет предобучаться сразу на таргетированной выборке (на которой потом и файнтьюним), что дает прирост в качестве
    c. актуальность для исследовательского сообщества
        i. думаю, все вышеперечисленные свойства делают статью актуальной для исследовательского сообщества. (вопрос лишь, насколько актуальной :/)
3. Слабые стороны: опишите недостатки статьи, следуя обозначенным выше критериям.
    a. В слабые стороны я бы отнесла часть с **воспроизводимостью результатов**

        i. об этом ниже
    b. и со **структурой повествования** статьи.
        i. несмотря на то, что статья подробная, почему-то нет описания методов с которыми осуществляется сравнение (BEiT, DINO) – просто дается ссылка на их статьи.
        ii. складывается впечатление, что первая часть работы (до splitMask)– подробное описание каких-то не оч важных вещей
    c. много утверждений дублируется (но мб это не минус, см пункт ниже)
4. Насколько хорошо написана статья: оцените насколько доходчиво написана статья, приведите примеры отрывков статьи, если такие есть, которые можно было бы доработать для улучшения восприятия статьи?
    a. статья написана доходчиво, это видно по ее объему; многие утверждения объясняются в разных вариациях
        i. This phenomenon is in- herent to pre-training with a fixed set of labels: the network learns to focus on the mapping between images and the la- bels of the pre-training stage, but can discard information that is relevant to other downstream tasks. ***In other terms,*** *pre-training on large-scale classification datasets does not necessarily align with the goal of learning general-purpose features, as it uses only a subset of the available information controlled by the given dataset categorization bias*
    b. хорошо объясняются графики и таблицы – и нет противоречия между тем что нарисовано и написано; они понятные
        i. drop is higher than using 10% ImageNet even though the numbers of samples is roughly the same. ***We hypothesis this is because*** *COCO images are not biased to be object-centric, while this joint embedding method was*
5. Воспроизводимость: статья достаточно подробна, чтобы можно было воспроизвести её основные результаты?
    a. ну вроде да, но я бы сказала что у них маловато кода, даже типа псевдокода нет
    b. Я бы сказала что воспроизвести можно, но это сложно
    c. С одной стороны приведены все детали конфигурации для воспроизведения статьи
    d. Но в то же время все приведенные детали ссылаются на множество других сторонних работы – и воспроизводимость тем самым усложняется. Т.е. секция с “Implementation details” – просто полный референс на другие работы
        “ we follow the pre- training hyperparameters of ***Bao et al.****”*
        “ We use the original ViT formulation as pro- posed by ***Dosovitskiy et al.****”*
        “ In order to obtain features compat- ible with the Feature Pyramid Network (FPN) design [69], we use max pooling and transposed convolution operations similar ***to El-Nouby et al”***
        “ We use the training hyper-parameters used by ***Liu et al****” (и тд)*
6. Поставьте оценку статье по десятибалльной шкале, следуя критериям ниже (критерии рецензии НИПСа).
    a. 8
7. Оцените вашу уверенность в оценке по пятибальной шкале, следуя критериям ниже (критерии уверенности НИПСа).
    a. 3 (оценка выше дилетантская и не уверена в ее адекватности :))( )