Mobile-PIRL: A Mobile Version of Self-Supervised Learning of Pretext-Invariant Representations

İzzet Emre Küçükkaya
Boğaziçi University
Istanbul, Turkey
izzet.kucukkaya@boun.edu.tr

Berkcan Üstün
Boğaziçi University
Istanbul, Turkey
berkcan.ustun@boun.edu.tr

***Abstract*—Supervised learning techniques of the Deep Neural Networks which are the main component of various applications in the domain of computer vision such as classification, segmentation or object detection are in need of the large amount of semantically annotated data in order to perform efficiently. However, it is a compelling situation to obtain a great quantity of semantically annotated data. The Self-Supervised learning techniques are introduced to overcome this obstacle. On the other hand, plenty of these techniques uses convolutional neural networks that have a great deal of complexity, parameters and inference time. In terms of performance, the recent mobile networks are considered to be comparable to complex network architectures despite their simple structure, few number parameters and less inference time. By this sense, the usage of newly introduced mobile networks in the domain of self supervised learning is an interesting topic of research. Regarding these facts, the performance of the mobile networks in a self supervised learning architecture can be examined. Our choice as the baseline architecture is PIRL (Pretext-Invariant Representation Learning). In addition, the number of transformations used in the PIRL can be increased along with the opportunity that they can be combined together.**

***Index Terms*—Mobile Networks, Data Augmentation, Self-Supervised Learning**

I. Introduction

Image representations are learned by modern image-recognition algorithms using massive datasets of photos and their semantic annotations. Class labels, hashtags, bounding boxes, and other types of annotations can be used to offer these annotations. Pre-defined semantic annotations do not scale well to the large tail of visual concepts, obstructing future image recognition advancements. Building more intelligent generalist models that can execute numerous applications and learn new abilities without vast quantities of labeled data is hampered by supervised learning. Self-supervised learning attempts to overcome these drawbacks by learning image representations from the pixels themselves rather than depending on those pre-defined lexical annotations. Many of the self-supervised learning techniques includes a pretext task which conducts a transformation to the input image with the need of the estimation of the properties of the transformation from the transformed image [1]–[3].

Although, often these techniques result on learning visual representations that are co-variant to the transformations, Pretext-Invariant Representation Learning (PIRL), which is one of the most successful methods, states the fact that

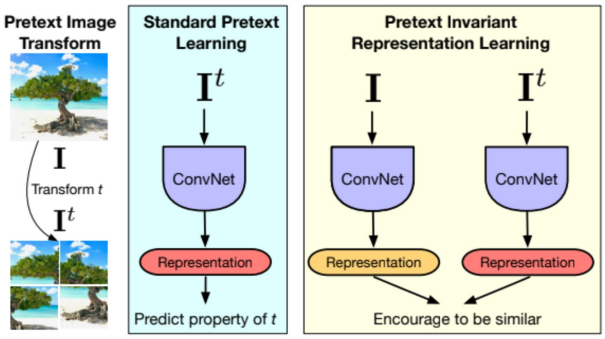

Fig. 1. Example of a figure caption.

learning co-variant representations is not useful for most of the visual recognition applications. In consideration of this, the PIRL underlines the necessity of learning invariant representations since the transformations do not change the semantics of the visual inputs. In order to obtain invariant representations, PIRL tries to create image representations that are close to other representations which are resulted in the transformation of the same image and far from the representations of other images. PIRL uses primarily the "Jigsaw" pretext task which is based on the division of the input image to nine patches and shuffling them. In addition, rotation pretext task presented along with the Jigsaw at the

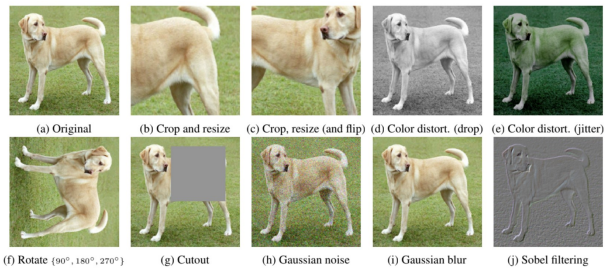

Fig. 2. Example of a figure caption.

II. Proposed Approach

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content

and organizational editing before formatting. Please note sections II-A–II-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads—LaTeX will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive".
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: "Wb/m^2" or "webers per square meter", not "webers/m^2". Spell out units when they appear in text: ". . . a few henries", not ". . . a few H".
- Use a zero before decimal points: "0.25", not ".25". Use "cm^3", not "cc".)

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{1}$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

D. LaTeX-Specific Advice

Please use "soft" (e.g., `\eqref{Eq}`) cross references instead of "hard" references (e.g., `(1)`). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don't use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you've discovered a new method of counting.

BibTeX does not work by magic. It doesn't get the bibliographic data from thin air but from .bib files. If you use BibTeX to produce a bibliography you must send the .bib files.

LaTeX can't read your mind. If you assign the same label to a subsubsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it's supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al.".
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [**?**].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 3", even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

Fig. 3. Example of a figure caption.

REFERENCES

[1] I. Misra and L. van der Maaten, "Self-Supervised Learning of Pretext-Invariant Representations," *arXiv e-prints*, p. arXiv:1912.01991, Dec. 2019.

[2] C. Doersch and A. Zisserman, "Multi-task Self-Supervised Visual Learning," *arXiv e-prints*, p. arXiv:1708.07860, Aug. 2017.

[3] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," *arXiv e-prints*, p. arXiv:2002.05709, Feb. 2020.