

# COSE474-2024F: Final Project Proposal

## “CLIP 기반 Zero-shot 학습에서 텍스트 프롬프트 설계의 영향 분석”

Youngmin Kim

### 1. Introduction

기존 연구(CoOp, CoCoOp 등)는 학습 가능한 프롬프트 최적화 방법에 초점을 맞추었지만, 단순한 프롬프트 설계가 CLIP의 성능에 미치는 영향을 체계적으로 분석한 연구는 부족합니다. 본 연구는 Zero-shot 환경에서 텍스트 프롬프트의 문구 및 구조 변화가 분류 성능에 미치는 영향을 탐구하며, 이를 통해 효과적인 프롬프트 설계에 대한 실질적인 통찰을 제공합니다.

### 2. Problem definition & challenges

본 연구의 목표는 다양한 프롬프트 스타일을 체계적으로 분석하고, 학습 과정 없이도 CLIP의 성능을 최대화할 수 있는 효과적인 프롬프트 설계 방안을 제시하는 것입니다.

이 문제를 해결하는 과정에서 다음과 같은 도전 과제에 직면합니다:

- 다양한 프롬프트 해석: CLIP은 텍스트 프롬프트의 미세한 차이에 민감하게 반응하여, 어떤 프롬프트가 최적의 성능을 낼지 예측하기 어렵습니다.
- 데이터셋 간 일반화: 특정 데이터셋에서 효과적인 프롬프트가 다른 데이터셋에서는 동일한 성능을 보장하지 않으며, 이는 이미지 스타일, 맥락, 또는 데이터셋 특유의 편향 때문입니다.
- 평가 복잡성: 단순한 정확도 측정만으로 텍스트와 이미지 간의 매칭 관계를 완전히 이해하기 어렵습니다. 보다 세부적인 분석과 평가 지표가 필요합니다.
- 표준화된 설계 가이드라인의 부재: 텍스트 프롬프트 설계를 위한 체계적인 가이드라인이 부족하며, 이를 탐구하는 연구가 필요합니다.

본 연구는 이러한 도전 과제를 해결하여 CLIP 기반 Zero-shot 학습의 실질적 활용성을 높이고, 텍스트 프롬프트 설계에 대한 새로운 통찰을 제공하고자 합니다.

### 3. Related Works

CLIP(Contrastive Language-Image Pretraining)은 Zero-shot 학습에서 강력한 성능을 보이며, 텍스트 프롬프트

설계가 성능에 중요한 역할을 한다는 점이 여러 연구에서 강조되었습니다. CoOp(Conditional Prompt Learning)과 CoCoOp(Context Optimization)은 CLIP의 프롬프트를 학습 가능한 파라미터로 최적화하여 성능을 향상시킨 대표적인 연구입니다. CoOp은 데이터셋에 맞춘 프롬프트 최적화를, CoCoOp은 문맥에 따른 동적인 프롬프트 조정을 가능하게 하여 범용성을 높였습니다. 그러나 이러한 접근법은 추가적인 학습 과정과 높은 계산 비용이 필요하며, CLIP의 Zero-shot 본연의 단순함을 유지하기 어렵습니다. 본 연구는 CoOp과 CoCoOp의 학습 기반 접근과 달리, 학습 없이 프롬프트 설계만으로 CLIP의 성능을 분석하고 최적화 가능성을 탐구합니다.

### 4. Datasets

본 연구에서는 CLIP의 Zero-shot 성능을 평가하기 위해 CIFAR-10 데이터셋을 사용하였습니다. CIFAR-10은 고양이, 개, 자동차 등 10개의 클래스와 6만 장의 이미지로 구성된 소규모 이미지 데이터셋으로, 각 클래스는 6,000개의 이미지를 포함합니다. 본 데이터셋은 이미지 크기가 32x32 픽셀로 고정되어 있으며, 간단하면서도 명확한 클래스 구분을 제공하여 Zero-shot 학습 실험에 적합합니다. 또한, 널리 사용되는 공개 데이터셋으로 연구 결과의 재현성을 높이고, 프롬프트 설계가 Zero-shot 성능에 미치는 영향을 분석하기에 효율적인 환경을 제공합니다.

### 5. State-of-the-art methods and baselines

CLIP은 Zero-shot 학습의 대표적인 모델로, 고정된 프롬프트 템플릿(예: "A photo of a [CLASS]")만으로도 높은 성능을 발휘합니다. 이를 개선하기 위해 CoOp(Conditional Prompt Learning)은 학습 가능한 템플릿을 사용하여 데이터셋에 최적화된 프롬프트를 생성하며, CoCoOp(Context Optimization)은 이를 확장해 도메인에 적응 가능한 동적 프롬프트를 제공합니다. 그러나 이러한 방법들은 추가적인 학습 과정과 높은 계산 비용을 요구합니다. 본 연구는 학습 기반 접근 대신, 프롬프트 설계만으로 CLIP의 Zero-shot 성능을 분석하고 최적화 가능성을 탐구하여 간단하면서도 효과적인 대안을 제시하고자 합니다.

### 6. Schedule & Roles

Research & Literature Review (2024-10-14 ~ 2024-10-20)

**Data Collection & Preprocessing (2024-10-21 ~2024-10-27)**

**Model Development (2024-10-28 ~2024-11-10)**

**Evaluation & Analysis (2024-11-11 ~2024-11-17)**

**Report Writing & Final Submission (2024-11-18 ~2024-11-24)**

The following is the author guideline of ICML. You can delete all.

Submission to ICML 2019 will be entirely electronic, via a web site (not email). Information about the submission process and L<sup>A</sup>T<sub>E</sub>X templates are available on the conference web site at:

<http://icml.cc/>

The guidelines below will be enforced for initial submissions and camera-ready copies. Here is a brief summary:

- Submissions must be in PDF.
- Submitted papers can be up to eight pages long, not including references, and up to twelve pages when references and acknowledgments are included. Any paper exceeding this length will automatically be rejected.
- **Do not include author information or acknowledgments** in your initial submission.
- Your paper should be in **10 point Times font**.
- Make sure your PDF file only uses Type-1 fonts.
- Place figure captions *under* the figure (and omit titles from inside the graphic file itself). Place table captions *over* the table.
- References must include page numbers whenever possible and be as complete as possible. Place multiple citations in chronological order.
- Do not alter the style template; in particular, do not compress the paper format by reducing the vertical spaces.
- Keep your abstract brief and self-contained, one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase. The title should have content words capitalized.

## 6.1. Submitting Papers

**Paper Deadline:** The deadline for paper submission that is advertised on the conference website is strict. If your full, anonymized, submission does not reach us on time, it will not be considered for publication.

**Anonymous Submission:** ICML uses double-blind review: no identifying author information may appear on the title page or in the paper itself. Section 7.3 gives further details.

**Simultaneous Submission:** ICML will not accept any paper which, at the time of submission, is under review for another conference or has already been published. This policy also applies to papers that overlap substantially in technical

content with conference papers under review or previously published. ICML submissions must not be submitted to other conferences during ICML’s review period. Authors may submit to ICML substantially different versions of journal papers that are currently under review by the journal, but not yet accepted at the time of submission. Informal publications, such as technical reports or papers in workshop proceedings which do not appear in print, do not fall under these restrictions.

Authors must provide their manuscripts in **PDF** format. Furthermore, please make sure that files contain only embedded Type-1 fonts (e.g., using the program `pdf fonts` in linux or using File/DocumentProperties/Fonts in Acrobat). Other fonts (like Type-3) might come from graphics files imported into the document.

Authors using **Word** must convert their document to PDF. Most of the latest versions of Word have the facility to do this automatically. Submissions will not be accepted in Word format or any format other than PDF. Really. We’re not joking. Don’t send Word.

Those who use L<sup>A</sup>T<sub>E</sub>X should avoid including Type-3 fonts. Those using `latex` and `dvips` may need the following two commands:

```
dvips -Ppdf -tletter -G0 -o paper.ps paper.dvi
ps2pdf paper.ps
```

It is a zero following the “-G”, which tells dvips to use the config.pdf file. Newer T<sub>E</sub>X distributions don’t always need this option.

Using `pdflatex` rather than `latex`, often gives better results. This program avoids the Type-3 font problem, and supports more advanced features in the `microtype` package.

**Graphics files** should be a reasonable size, and included from an appropriate format. Use vector formats (.eps/.pdf) for plots, lossless bitmap formats (.png) for raster graphics with sharp lines, and jpeg for photo-like images.

The style file uses the `hyperref` package to make clickable links in documents. If this causes problems for you, add `nohyperref` as one of the options to the `icml2019` usepackage statement.

## 6.2. Submitting Final Camera-Ready Copy

The final versions of papers accepted for publication should follow the same format and naming convention as initial submissions, except that author information (names and affiliations) should be given. See Section 7.3.2 for formatting instructions.

The footnote, “Preliminary work. Under review by the In-

ternational Conference on Machine Learning (ICML). Do not distribute.” must be modified to “*Proceedings of the 36<sup>th</sup> International Conference on Machine Learning*, Long Beach, USA, 2019. Copyright 2019 by the author(s).”

For those using the **L<sup>A</sup>T<sub>E</sub>X** style file, this change (and others) is handled automatically by simply changing `\usepackage{icml2019}` to

```
\usepackage[accepted]{icml2019}
```

Authors using **Word** must edit the footnote on the first page of the document themselves.

Camera-ready copies should have the title of the paper as running head on each page except the first one. The running title consists of a single line centered above a horizontal rule which is 1 point thick. The running head should be centered, bold and in 9 point type. The rule should be 10 points above the main text. For those using the **L<sup>A</sup>T<sub>E</sub>X** style file, the original title is automatically set as running head using the `fancyhdr` package which is included in the ICML 2019 style file package. In case that the original title exceeds the size restrictions, a shorter form can be supplied by using

```
\icmltitlerunning{...}
```

just before `\begin{document}`. Authors using **Word** must edit the header of the document themselves.

## 7. Format of the Paper

All submissions must follow the specified format.

### 7.1. Length and Dimensions

Submitted papers can be up to eight pages long, not including references, and up to twelve pages when references and acknowledgments are included. Acknowledgments should be limited to grants and people who contributed to the paper. Any submission that exceeds this page limit, or that diverges significantly from the specified format, will be rejected without review.

The text of the paper should be formatted in two columns, with an overall width of 6.75 inches, height of 9.0 inches, and 0.25 inches between the columns. The left margin should be 0.75 inches and the top margin 1.0 inch (2.54 cm). The right and bottom margins will depend on whether you print on US letter or A4 paper, but all final versions must be produced for US letter size.

The paper body should be set in 10 point type with a vertical spacing of 11 points. Please use Times typeface throughout the text.

### 7.2. Title

The paper title should be set in 14 point bold type and centered between two horizontal rules that are 1 point thick, with 1.0 inch between the top rule and the top edge of the page. Capitalize the first letter of content words and put the rest of the title in lower case.

### 7.3. Author Information for Submission

ICML uses double-blind review, so author information must not appear. If you are using **L<sup>A</sup>T<sub>E</sub>X** and the `icml2019.sty` file, use `\icmlauthor{...}` to specify authors and `\icmlaffiliation{...}` to specify affiliations. (Read the TeX code used to produce this document for an example usage.) The author information will not be printed unless `accepted` is passed as an argument to the style file. Submissions that include the author information will not be reviewed.

#### 7.3.1. SELF-CITATIONS

If you are citing published papers for which you are an author, refer to yourself in the third person. In particular, do not use phrases that reveal your identity (e.g., “in previous work (?), we have shown ...”).

Do not anonymize citations in the reference section. The only exception are manuscripts that are not yet published (e.g., under submission). If you choose to refer to such unpublished manuscripts (?), anonymized copies have to be submitted as Supplementary Material via CMT. However, keep in mind that an ICML paper should be self contained and should contain sufficient detail for the reviewers to evaluate the work. In particular, reviewers are not required to look at the Supplementary Material when writing their review.

#### 7.3.2. CAMERA-READY AUTHOR INFORMATION

If a paper is accepted, a final camera-ready copy must be prepared. For camera-ready papers, author information should start 0.3 inches below the bottom rule surrounding the title. The authors’ names should appear in 10 point bold type, in a row, separated by white space, and centered. Author names should not be broken across lines. Unbolded superscripted numbers, starting 1, should be used to refer to affiliations.

Affiliations should be numbered in the order of appearance. A single footnote block of text should be used to list all the affiliations. (Academic affiliations should list Department, University, City, State/Region, Country. Similarly for industrial affiliations.)

Each distinct affiliations should be listed once. If an author has multiple affiliations, multiple superscripts should be placed after the name, separated by thin spaces. If the au-

thors would like to highlight equal contribution by multiple first authors, those authors should have an asterisk placed after their name in superscript, and the term “\*Equal contribution” should be placed in the footnote block ahead of the list of affiliations. A list of corresponding authors and their emails (in the format Full Name <email@domain.com>) can follow the list of affiliations. Ideally only one or two names should be listed.

A sample file with author names is included in the ICML2019 style file package. Turn on the `[accepted]` option to the stylefile to see the names rendered. All of the guidelines above are implemented by the `LATEX` style file.

#### 7.4. Abstract

The paper abstract should begin in the left column, 0.4 inches below the final address. The heading ‘Abstract’ should be centered, bold, and in 11 point type. The abstract body should use 10 point type, with a vertical spacing of 11 points, and should be indented 0.25 inches more than normal on left-hand and right-hand margins. Insert 0.4 inches of blank space after the body. Keep your abstract brief and self-contained, limiting it to one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase.

#### 7.5. Partitioning the Text

You should organize your paper into sections and paragraphs to help readers place a structure on the material and understand its contributions.

##### 7.5.1. SECTIONS AND SUBSECTIONS

Section headings should be numbered, flush left, and set in 11 pt bold type with the content words capitalized. Leave 0.25 inches of space before the heading and 0.15 inches after the heading.

Similarly, subsection headings should be numbered, flush left, and set in 10 pt bold type with the content words capitalized. Leave 0.2 inches of space before the heading and 0.13 inches afterward.

Finally, subsubsection headings should be numbered, flush left, and set in 10 pt small caps with the content words capitalized. Leave 0.18 inches of space before the heading and 0.1 inches after the heading.

Please use no more than three levels of headings.

##### 7.5.2. PARAGRAPHS AND FOOTNOTES

Within each section or subsection, you should further partition the paper into paragraphs. Do not indent the first line of a given paragraph, but insert a blank line between succeed-

ing ones.

You can use footnotes<sup>1</sup> to provide readers with additional information about a topic without interrupting the flow of the paper. Indicate footnotes with a number in the text where the point is most relevant. Place the footnote in 9 point type at the bottom of the column in which it appears. Precede the first footnote in a column with a horizontal rule of 0.8 inches.<sup>2</sup>

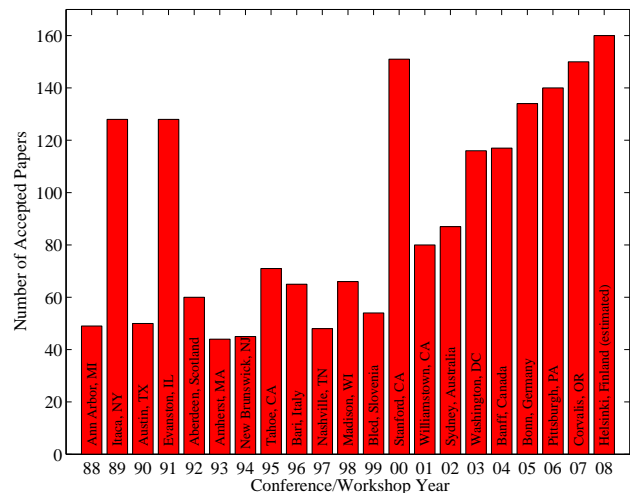


Figure 1. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead estimated.

#### 7.6. Figures

You may want to include figures in the paper to illustrate your approach and results. Such artwork should be centered, legible, and separated from the text. Lines should be dark and at least 0.5 points thick for purposes of reproduction, and text should not appear on a gray background.

Label all distinct components of each figure. If the figure takes the form of a graph, then give a name for each axis and include a legend that briefly describes each curve. Do not include a title inside the figure; instead, the caption should serve this function.

Number figures sequentially, placing the figure number and caption *after* the graphics, with at least 0.1 inches of space before the caption and 0.1 inches after it, as in Figure 1. The

<sup>1</sup>Footnotes should be complete sentences.

<sup>2</sup>Multiple footnotes can appear in each column, in the same order as they appear in the text, but spread them across columns and pages if possible.

**Algorithm 1** Bubble Sort

---

**Input:** data  $x_i$ , size  $m$

**repeat**

  Initialize  $noChange = true$ .

**for**  $i = 1$  **to**  $m - 1$  **do**

**if**  $x_i > x_{i+1}$  **then**

      Swap  $x_i$  and  $x_{i+1}$

$noChange = false$

**end if**

**end for**

**until**  $noChange$  is  $true$

---

Table 1. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

DATA SET	NAIVE	FLEXIBLE	BETTER?
BREAST	95.9 ± 0.2	96.7 ± 0.2	✓
CLEVELAND	83.3 ± 0.6	80.0 ± 0.6	×
GLASS2	61.9 ± 1.4	83.8 ± 0.7	✓
CREDIT	74.8 ± 0.5	78.3 ± 0.6	
HORSE	73.3 ± 0.9	69.7 ± 1.0	×
META	67.1 ± 0.6	76.5 ± 0.5	✓
PIMA	75.1 ± 0.6	73.9 ± 0.5	
VEHICLE	44.9 ± 0.6	61.5 ± 0.4	✓

figure caption should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left. You may float figures to the top or bottom of a column, and you may set wide figures across both columns (use the environment `figure*` in  $\LaTeX$ ). Always place two-column figures at the top or bottom of the page.

## 7.7. Algorithms

If you are using  $\LaTeX$ , please use the “algorithm” and “algorithmic” environments to format pseudocode. These require the corresponding stylefiles, `algorithm.sty` and `algorithmic.sty`, which are supplied with this package. Algorithm 1 shows an example.

## 7.8. Tables

You may also want to include tables that summarize material. Like figures, these should be centered, legible, and numbered consecutively. However, place the title *above* the table with at least 0.1 inches of space before the title and the same after it, as in Table 1. The table title should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left.

Tables contain textual material, whereas figures contain graphical material. Specify the contents of each row and column in the table’s topmost row. Again, you may float tables to a column’s top or bottom, and set wide tables across

both columns. Place two-column tables at the top or bottom of the page.

## 7.9. Citations and References

Please use APA reference format regardless of your formatter or word processor. If you rely on the  $\LaTeX$  bibliographic facility, use `natbib.sty` and `icml2019.bst` included in the style-file package to obtain this format.

Citations within the text should include the authors’ last names and year. If the authors’ names are included in the sentence, place only the year in parentheses, for example when referencing Arthur Samuel’s pioneering work (?). Otherwise place the entire reference in parentheses with the authors and year separated by a comma (?). List multiple references separated by semicolons (???). Use the ‘et al.’ construct only for citations with three or more authors or after listing all authors to a publication in an earlier reference (?).

Authors should cite their own work in the third person in the initial version of their paper submitted for blind review. Please refer to Section 7.3 for detailed instructions on how to cite your own papers.

Use an unnumbered first-level section heading for the references, and use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 10 points. The references at the end of this document give examples for journal articles (?), conference publications (?), book chapters (?), books (?), edited volumes (?), technical reports (?), and dissertations (?).

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

Please put some effort into making references complete, presentable, and consistent. If using `bibtex`, please protect capital letters of names and abbreviations in titles, for example, use `{B}ayesian` or `{L}ipschitz` in your `.bib` file.

## 7.10. Software and Data

We strongly encourage the publication of software and data with the camera-ready version of the paper whenever appropriate. This can be done by including a URL in the camera-ready copy. However, do not include URLs that reveal your institution or identity in your submission for review. Instead, provide an anonymous URL or upload the material as “Supplementary Material” into the CMT reviewing system. Note that reviewers are not required to look at this material when writing their review.

## Acknowledgements

**Do not** include acknowledgements in the initial version of the paper submitted for blind review.

If a paper is accepted, the final camera-ready version can (and probably should) include acknowledgements. In this case, please place such acknowledgements in an unnumbered section at the end of the paper. Typically, this will include thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.

### A. *Do not* have an appendix here

***Do not put content after the references.*** Put anything that you might normally include after the references in a separate supplementary file.

We recommend that you build supplementary material in a separate document. If you must create one PDF and cut it up, please be careful to use a tool that doesn't alter the margins, and that doesn't aggressively rewrite the PDF file. pdftk usually works fine.

**Please do not use Apple's preview to cut off supplementary material.** In previous years it has altered margins, and created headaches at the camera-ready stage.