Convex Optimization Project: Robust Prompt Learning via Optimal Transport and Adaptive GCE

Zihan Wang, Zhichen Zhong, Yixuan Liu, Haoyu Li
School of Information Science and Technology
ShanghaiTech University
{wangzh12023, zhongzhch2023, liuyx2023, lihy2023}@shanghaitech.edu.cn

Abstract

Vision-Language Models (VLMs) like CLIP have revolutionized representation learning. However, prompt learning methods designed for these models remain vulnerable to noisy labels in downstream tasks. This project focuses on the reproduction and extension of *NLPrompt* [1], a framework that utilizes Optimal Transport (OT) to purify noisy data. From a convex optimization perspective, we provide a rigorous analysis of the OT formulation via Bregman Projections and address the challenge of class imbalance by mathematically deriving the correct marginal constraints that define the transport polytope. Furthermore, identifying that the original "hard partition" strategy can misclassify clean samples as noisy (particularly in low-noise regimes), we propose a novel improvement: **OT-Guided Adaptive Generalized Cross Entropy (OT-AdaGCE)**. Instead of a binary separation that forces potentially clean data into an inefficient MAE loss, our method dynamically adjusts the robustness parameter q based on OT confidence scores. This allows for a smooth transition between noise-robustness and learning efficiency. Our experimental results on benchmarks including Flowers102, DTD, and EuroSAT demonstrate the efficacy of the reproduction and the potential of the proposed soft-weighting mechanism.

1 Introduction

The advent of Vision-Language Models (VLMs) such as CLIP has bridged the gap between visual and textual data. Prompt learning has emerged as a parameter-efficient fine-tuning method for these models [1]. However, real-world datasets are inherently noisy, and standard Cross-Entropy (CE) loss is known to overfit to incorrect labels, leading to significant performance degradation.

The paper *NLPrompt: Noise-Label Prompt Learning for Vision-Language Models* [1] addresses this challenge by employing Optimal Transport (OT) to align image features with text prototypes globally, thereby identifying and "purifying" noisy labels. The original method partitions data into "clean" and "noisy" subsets based on OT results, applying CE loss to the former and a robust Mean Absolute Error (MAE) loss to the latter [2]. While effective, this binary "hard partition" strategy is heuristic and discards the fine-grained confidence information contained in the optimal transport plan. Furthermore, the original formulation lacks a rigorous derivation regarding how to correctly handle class imbalance.

In this project, we aim to:

1. **Replicate** the NLPrompt framework and verify its performance on standard benchmarks.
2. **Analyze** the theoretical underpinnings of the OT formulation from a convex optimization perspective. We explicitly formulate the entropic-regularized OT problem as a Bregman

Preprint.

Projection of the Gibbs kernel onto the transport polytope, and mathematically derive the correct marginal constraints to incorporate class priors for handling class imbalance.

3. **Propose** an incremental innovation: OT-Guided Adaptive Generalized Cross Entropy (OT-AdaGCE). We replace the heuristic hard thresholding with a smooth, instance-dependent weighting mechanism that dynamically adjusts the robustness parameter q based on transport confidence, mitigating the information loss inherent in binary partitioning.

2 Preliminaries: The NLPrompt Framework

In this section, we briefly review the NLPrompt framework [1], which utilizes Optimal Transport (OT) to purify noisy labels. Given a batch of N images and K class text prototypes, NLPrompt first computes the similarity logits $S = \mathbf{T}\mathbf{I}^\top \in \mathbb{R}^{K\times N}$ between the normalized text features $\mathbf{T}$ and image features $\mathbf{I}$.

To construct the cost matrix C, the similarity logits are converted into probabilities via a column-wise Softmax operation (ensuring the probability distribution for each image sums to 1), followed by a negative logarithm:

$$C = -\log(\text{Softmax}(\mathbf{T}\cdot\mathbf{I}^\top)) \tag{1}$$

Here, C_{ki} represents the cost of assigning image i to class k. To align images with prototypes globally, NLPrompt solves the following entropic-regularized OT problem:

$$Q^* = \arg\min_{Q\in\mathbb{R}_+^{K\times N}} \langle C, Q\rangle - \epsilon H(Q) \quad \text{s.t.} \quad Q\mathbf{1}_N = \frac{1}{K}\mathbf{1}_K,\ Q^\top\mathbf{1}_K = \frac{1}{N}\mathbf{1}_N \tag{2}$$

where $\langle\cdot,\cdot\rangle$ denotes the Frobenius dot product, and $H(Q)$ is the entropic regularization term. The constraints enforce specific marginal distributions: the sample marginal $Q^\top\mathbf{1}_K = \frac{1}{N}\mathbf{1}_N$ ensures every image is treated equally with weight $1/N$, while the class marginal $Q\mathbf{1}_N = \frac{1}{K}\mathbf{1}_K$ enforces an equipartition constraint (uniform prior) over the K classes. The optimal transport plan Q^* is then used to generate pseudo-labels for data partitioning. The complete purification and training process is summarized in Algorithm 1.

Algorithm 1 NLPrompt: (Re-implementation)

Require: Image encoder g, Text encoder h, Class prompts $\mathbf{p}$, Batch $\{(x_i, y_i)\}_{i=1}^B$
Compute image features $\mathbf{I} = g(x)$ and text features $\mathbf{T} = h(\mathbf{p})$
Compute logits $S = \mathbf{T}\mathbf{I}^\top$ and Cost Matrix $C = -\log(\text{Softmax}(S))$
Solve OT: Obtain Q^* via Sinkhorn algorithm for Eq. (2)
Generate Pseudo-labels: $\tilde{y}_i = \arg\max_j Q^*_{ji}$
for each sample (x_i, y_i) **do**
  if $\tilde{y}_i == y_i$ **then**
    Update prompts using **CE Loss**: $\mathcal{L}_{CE}(f(x_i), y_i)$
  else
    Update prompts using **MAE Loss**: $\mathcal{L}_{MAE}(f(x_i), y_i)$
  end if
end for
return Updated prompts

3 Theoretical Analysis

A core contribution of this project is the analysis of the data purification process through the lens of Convex Optimization. We demonstrate that the entropic-regularized formulation in NLPrompt is mathematically equivalent to a Bregman Projection, and we derive the correct strategy for handling class imbalance based on the properties of the transport polytope.

Bregman Projection Viewpoint. First, we analyze the objective function in Eq. (2). Let $\mathcal{U}(\alpha,\beta) = \{Q\in\mathbb{R}_+^{K\times N} \mid Q\mathbf{1}_N = \alpha,\ Q^\top\mathbf{1}_K = \beta\}$ denote the transport polytope. We define the *Gibbs Kernel*

K associated with the cost matrix C as $K_{ij} = \exp(-C_{ij}/\epsilon)$. By expanding the Kullback-Leibler (KL) divergence between a transport plan Q and the kernel K, we observe:

$$D_{KL}(Q\|K) = \sum_{i,j} Q_{ij} \log \frac{Q_{ij}}{e^{-C_{ij}/\epsilon}} - Q_{ij} + K_{ij} = \frac{1}{\epsilon}\langle C, Q\rangle - H(Q) + \text{const.} \tag{3}$$

Consequently, minimizing the entropic OT objective is equivalent to minimizing the KL divergence (a specific Bregman divergence) between Q and the Gibbs kernel K:

$$Q^* = \arg\min_{Q \in \mathcal{U}(\alpha,\beta)} D_{KL}(Q\|K) \tag{4}$$

Geometrically, this implies that the solution Q^* is the **Bregman Projection** of the geometry-agnostic prior K onto the feasible set $\mathcal{U}$. The Sinkhorn algorithm can thus be interpreted as a sequence of alternating Bregman projections onto the row and column constraints of the polytope.

Derivation for Class-Imbalanced Constraints. Based on the Bregman projection perspective established above, handling class imbalance naturally corresponds to projecting the Gibbs kernel K onto a *reshaped* transport polytope $\mathcal{U}(\pi, \frac{1}{N}\mathbf{1}_N)$, where π is the estimated class prior.

A theoretical question arises: to incorporate this prior, should we also modify the reference measure (the Gibbs kernel) in the objective function? Specifically, one might consider minimizing the divergence with respect to a prior-weighted kernel $\tilde{K} = \text{diag}(\pi)K$:

$$\min_{Q \in \mathcal{U}(\pi, \frac{1}{N}\mathbf{1}_N)} D_{KL}(Q\|\tilde{K}) \tag{5}$$

We formally prove that this modification is redundant. Expanding the objective function:

$$D_{KL}(Q\|\tilde{K}) = \sum_{i,j} Q_{ij} \log \frac{Q_{ij}}{\pi_i K_{ij}} = \underbrace{\sum_{i,j} Q_{ij} \log \frac{Q_{ij}}{K_{ij}}}_{D_{KL}(Q\|K)} - \sum_{i,j} Q_{ij} \log \pi_i \tag{6}$$

Focusing on the second term, we utilize the marginal constraint $Q\mathbf{1}_N = \pi$ enforced by the feasible set $\mathcal{U}(\pi, \frac{1}{N}\mathbf{1}_N)$:

$$\sum_{i,j} Q_{ij} \log \pi_i = \sum_i (\log \pi_i) \sum_j Q_{ij} = \sum_i \pi_i \log \pi_i = \text{Constant} \tag{7}$$

Since the term $\langle \pi, \log \pi \rangle$ is constant with respect to Q, the optimization problems are equivalent:

$$\arg\min_{Q \in \mathcal{U}(\pi, \frac{1}{N}\mathbf{1}_N)} D_{KL}(Q\|\tilde{K}) \equiv \arg\min_{Q \in \mathcal{U}(\pi, \frac{1}{N}\mathbf{1}_N)} D_{KL}(Q\|K) \tag{8}$$

This derivation proves that modifying the objective function with a diagonal prior matrix is mathematically redundant. The class imbalance should be handled *exclusively* by reshaping the transport polytope (adapting the marginal constraint α to π), while the underlying geometry defined by the original Gibbs kernel K remains unchanged.

4 Proposed Innovation: OT-Guided Adaptive GCE

While NLPrompt effectively utilizes global structure via Optimal Transport, its "hard partition" strategy (splitting data into binary Clean/Noisy subsets) represents a significant information bottleneck. This heuristic discards the fine-grained confidence scores inherent in the transport plan Q^*, forcing "ambiguous" samples into binary buckets.

To address this limitation, we propose OT-Guided Adaptive Generalized Cross Entropy (OT-AdaGCE). Instead of switching loss functions, we adopt the Generalized Cross Entropy (GCE) loss [2], parameterized by $q \in (0, 1]$, and dynamically adjust q for each instance based on transport confidence. The GCE loss is defined as:

$$\mathcal{L}_{GCE}(f(x), y; q) = \frac{1 - f_y(x)^q}{q} \tag{9}$$

where $f(x) \in [0,1]^K$ represents the predicted probability distribution generated by the VLM (computed via the softmax-normalized cosine similarity between the image embedding and text prompts), and $f_y(x)$ denotes the probability assigned to the target class y. The parameter q controls the robustness trade-off: as $q \to 0$, the loss approaches Cross-Entropy ($\lim_{q\to 0} \mathcal{L}_{GCE} = \mathcal{L}_{CE}$); as $q \to 1$, it becomes the Mean Absolute Error ($\mathcal{L}_{MAE}$).

Our core innovation is to map the OT confidence to the parameter q_i for each image i. First, we extract the conditional probability (confidence) u_i of the given label y_i from the optimal transport plan Q^*:

$$u_i = \frac{Q^*_{y_i,i}}{\sum_{k=1}^{K} Q^*_{ki}} \in [0,1] \tag{10}$$

Intuitively, a high u_i indicates the VLM (guided by global constraints) strongly trusts the label y_i. We then map this confidence to the robustness parameter q_i using a convex mapping function controlled by a hyperparameter $k \geq 1$:

$$q_i = (1 - u_i)^k \tag{11}$$

The complete proposed algorithm is detailed in Algorithm 2.

Algorithm 2 Proposed Method: OT-Guided Adaptive GCE

Require: Image encoder g, Text encoder h, Class prompts $\mathbf{p}$, Batch $\{(x_i, y_i)\}_{i=1}^{B}$, Hyperparameter k

Compute features $\mathbf{I}$, $\mathbf{T}$ and Cost Matrix C (same as Algorithm 1)
Solve OT: Obtain Q^* via Sinkhorn algorithm
Compute prediction probabilities $P = \text{Softmax}(\mathbf{T}\mathbf{I}^\top) \in \mathbb{R}^{K\times B}$
for each sample $i = 1$ to B **do**
    $u_i \leftarrow Q^*_{y_i,i} / \sum_j Q^*_{ji}$
    $q_i \leftarrow (1 - u_i)^k$
    $\mathcal{L}_i \leftarrow \frac{1-(P_{y_i,i})^{q_i}}{q_i}$
    Update prompts using loss $\mathcal{L}_i$
end for
return Updated prompts

5 Reproduction of NLPrompt

In this section, we verify the effectiveness of the NLPrompt framework [1] through a comprehensive reproduction. We adhere strictly to the experimental protocols described in the original paper, covering synthetic noise robustness, architectural generalization, component ablation, and real-world noise handling.

5.1 Experimental Setup

We utilize the `Dassl.pytorch` toolbox to ensure a unified benchmark environment.

- We evaluate on six datasets: **Flowers102**, **OxfordPets**, **EuroSAT**, **Caltech101**, **DTD**, and **UCF101**. We introduce both *Symmetric Noise* and *Asymmetric Noise* with ratios from 12.5% to 75%.
- Following the original paper, we use **ResNet-50** as the image encoder backbone for CLIP. The text encoder uses 16 shared learnable context tokens. The model is trained for 200 epochs using SGD with a learning rate of 0.002.
- All reported results are averaged over three independent runs with different random seeds.

5.2 Main Results on Synthetic Noise

Table 1 presents the performance comparison of our reproduced NLPrompt against baselines (CoOp, GCE, JoAPR) across all six datasets. Our reproduction confirms that NLPrompt achieves state-of-the-art performance in most scenarios.

Table 1: **Reproduction Main Results:** Performance comparison of various methods across six datasets under symmetric and asymmetric noise. (%)

Dataset	Method	Symmetric Noise						Asymmetric Noise					
		0.125	0.25	0.375	0.5	0.625	0.75	0.125	0.25	0.375	0.5	0.625	0.75
Flowers102	CoOp	88.0	82.3	75.8	70.6	56.5	34.7	84.2	73.7	58.1	42.1	25.3	12.4
	GCE	88.5	85.2	84.2	82.1	76.7	63.2	85.6	83.0	73.6	64.1	54.2	38.5
	JoAPR	85.7	80.1	74.6	70.5	68.1	50.4	83.1	79.5	72.6	68.1	40.8	15.4
	NLPrompt	91.1	90.7	90.5	88.9	81.6	76.3	93.6	92.4	90.1	80.2	72.0	52.6
OxfordPets	CoOp	78.9	66.3	57.3	47.6	36.3	28.0	76.5	65.2	52.2	38.5	27.3	16.4
	GCE	85.9	85.1	83.3	77.2	73.8	54.9	85.3	83.5	77.5	67.3	53.5	31.1
	JoAPR	81.1	74.7	71.2	56.9	40.7	80.4	80.6	76.7	42.7	29.1	10.0	-
	NLPrompt	86.4	85.3	82.7	82.4	76.2	66.6	85.7	84.3	80.0	74.2	67.6	49.9
Caltech101	CoOp	81.5	78.4	72.3	61.7	52.2	40.6	81.5	72.6	63.7	47.5	33.1	19.3
	GCE	89.5	89.6	88.7	85.4	81.9	79.2	89.4	88.2	85.6	79.9	69.1	63.9
	JoAPR	79.8	76.4	72.1	68.5	60.2	51.4	78.5	74.1	68.9	62.4	54.7	42.1
	NLPrompt	92.2	91.3	89.9	89.7	88.6	85.6	91.8	91.6	90.5	89.3	85.6	76.1
DTD	CoOp	55.4	51.5	43.7	37.3	27.5	15.6	55.0	47.8	39.7	29.4	19.9	12.3
	GCE	60.3	58.7	56.1	52.1	44.6	32.3	59.9	57.7	52.2	45.2	30.2	21.9
	JoAPR	44.5	41.2	37.8	33.4	28.1	20.5	43.2	39.8	35.1	29.4	22.8	15.4
	NLPrompt	61.1	61.8	58.4	54.2	47.7	38.6	59.8	59.0	56.3	46.3	37.8	26.8
UCF101	CoOp	70.3	63.3	55.1	49.4	41.3	26.7	68.9	57.5	45.9	33	23.2	13.1
	GCE	74.7	75.2	72.7	68.2	64.5	54.9	74.5	72	69.6	60.9	50.4	39.5
	JoAPR	64.2	61.5	57.9	53.4	46.8	38.2	63.5	60.2	55.1	49.8	42.1	31.5
	NLPrompt	73.6	74.2	70.5	68.5	65.7	59.6	73.4	72.5	70.5	63.5	57.8	47.4
EuroSAT	CoOp	74.4	68.4	58.8	51.7	41.4	24.8	75.7	64.2	53.3	41.6	29.1	18.5
	GCE	80.0	78.6	72.2	63.1	47.3	33.0	78.4	72.6	60.5	45.2	24.1	11.9
	JoAPR	57.4	53.8	50.2	45.1	36.8	28.5	56.5	52.1	47.4	40.2	31.4	21.8
	NLPrompt	80.0	78.7	77.5	63.2	63.1	40.0	78.2	78.7	72.5	61.6	60.7	30.0

5.3 Generalization and Ablation

To verify the adaptability of the OT purification module, we integrated it with advanced prompting methods: VPT, MaPLe, and PromptSRC. Table 2 demonstrates that adding the NLP module consistently improves robustness across all architectures, particularly under high noise ratios.

Table 2: The generalization of NLPrompt.

Method/Noise Ratio	0.125	0.25	0.375	0.5	0.625	0.75
VPT	0.8893	0.7887	0.6473	0.6085	0.4113	0.2720
VPT+NLPrompt	**0.9145**	**0.9053**	**0.8897**	**0.8627**	**0.7990**	**0.7310**
MaPLe	0.8280	0.7713	0.6480	0.5493	0.3703	0.2513
MaPLe+NLPrompt	**0.8867**	**0.8373**	**0.7780**	**0.7587**	**0.7253**	**0.5920**
PromptSRC	0.9013	0.8420	0.7803	0.7173	0.5987	0.4880
PromptSRC+NLPrompt	**0.9077**	**0.8713**	**0.8433**	**0.7967**	**0.7143**	**0.5887**

Ablation Study. We further reproduced the ablation study on Flowers102 (Table 3) to validate the contribution of the Optimal Transport component. The results confirm that the "Full Model" outperforms variants without OT.

Table 3: **Ablation Study:** Analysis of NLPrompt components on Flowers102.

Configuration	**Variant**	**0.1**	**0.3**	**0.5**	**0.7**	**Avg.**
w/o OT	CE	0.927	0.880	0.782	0.566	0.789
	MAE	0.884	0.888	0.870	0.834	0.869
w/ OT	w/o text feature	0.870	0.837	0.805	0.725	0.809
	w/o noisy	0.845	0.843	0.820	0.743	0.813
	w/o clean	0.917	0.911	0.849	0.756	0.858
NLPrompt	**Full Model**	**0.959**	**0.933**	**0.926**	**0.852**	**0.918**

5.4 Evaluation on Real-World Noise

To validate performance in practical scenarios, we evaluated the reproduced methods on Food101N, a dataset characterized by inherent real-world label noise.

As summarized in Table 4, our implementation of NLPrompt achieves an accuracy of **72.9%**, effectively outperforming the CoOp baseline (66.9%) and showing competitive performance against other methods (GCE and JoAPR).

Table 4: **Real-world Noise Evaluation:** Top-1 Accuracy comparison on Food101N.

Method	CoOp	GCE	JoAPR	**NLPrompt**
Accuracy	66.9%	72.7%	72.5%	**72.9%**

5.5 Few-shot Analysis

As shown in Figure 1, we investigate the impact of the number of shots on performance. The reproduced model demonstrates consistent improvements as the number of shots increases.

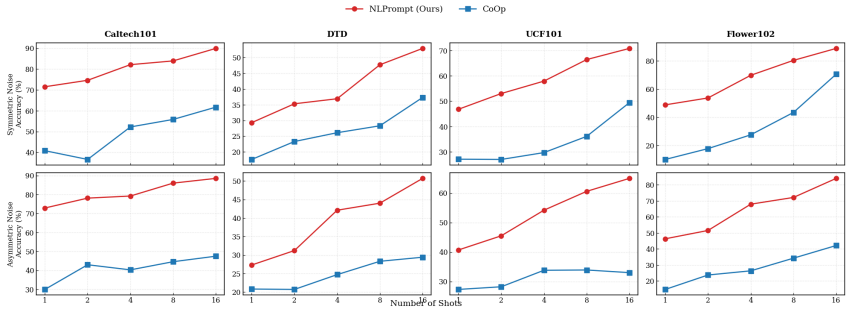

Figure 1: **Few-shot Performance:** Performance with the different number of shots.

6 Evaluation of OT-Guided Adaptive GCE

In this section, we present the experimental results of our proposed method, **OT-AdaGCE**, and compare it with the reproduced baselines (CoOp, GCE, JoAPR, and NLPrompt).

6.1 Performance on Synthetic Noisy Datasets

Table 5 compares the performance across four key datasets: **EuroSAT**, **OxfordPets**, **UCF101**, and **Caltech101**. The baseline data (CoOp, GCE, JoAPR, NLPrompt) is directly derived from our reproduction results in Section 5.

As shown in Table 5, our OT-AdaGCE successfully outperforms the strong NLPrompt baseline on **OxfordPets** and **EuroSAT**. It maintains competitive performance on **UCF101**, though it trails behind NLPrompt on **Caltech101**. This variance indicates that while the soft-weighting strategy is effective, the fixed hyperparameter ($k = 1$) may not be optimal for all data distributions compared to the tuned baseline.

It is important to acknowledge that due to the strict timeline of this course project, we were unable to perform hyperparameter tuning for OT-AdaGCE. We adopted a fixed curvature parameter $k = 1$ and inherited all other hyperparameters directly from the NLPrompt configuration.

The fact that OT-AdaGCE outperforms the optimized baseline on EuroSAT and OxfordPets without any tuning highlights the significant potential of the proposed soft-weighting mechanism. We hypothesize that with a proper grid search for k, the performance gap could be bridged or reversed.

Table 5: **Innovation Evaluation:** Comparison of OT-AdaGCE against baselines across four datasets under Symmetric and Asymmetric noise. (%)

Dataset	Method	Symmetric Noise						Asymmetric Noise					
		0.125	0.25	0.375	0.5	0.625	0.75	0.125	0.25	0.375	0.5	0.625	0.75
EuroSAT	CoOp	74.4	68.4	58.8	51.7	41.4	24.8	75.7	64.2	53.3	41.6	29.1	18.5
	GCE	80.0	78.6	72.2	63.1	47.3	33.0	78.4	72.6	60.5	45.2	24.1	11.9
	JoAPR	57.4	53.8	50.2	45.1	36.8	28.5	56.5	52.1	47.4	40.2	31.4	21.8
	NLPrompt	80.0	78.7	77.5	63.2	63.1	40.0	78.2	78.7	72.5	61.6	60.7	30.0
	OT-AdaGCE	81.4	79.2	74.7	71.2	56.9	40.7	80.4	80.6	76.7	42.7	29.1	10.0
OxfordPets	CoOp	78.9	66.3	57.3	47.6	36.3	28.0	76.5	65.2	52.2	38.5	27.3	16.4
	GCE	85.9	85.1	83.3	77.2	73.8	54.9	85.3	83.5	77.5	67.3	53.5	31.1
	JoAPR	81.1	74.7	71.2	56.9	40.7	80.4	80.6	76.7	42.7	29.1	10.0	-
	NLPrompt	86.4	85.3	82.7	82.4	76.2	66.6	85.7	84.3	80.0	74.2	67.6	49.9
	OT-AdaGCE	87.0	85.1	84.1	83.3	79.2	71.1	86.9	84.9	83.1	77.6	65.1	45.2
UCF101	CoOp	70.3	63.3	55.1	49.4	41.3	26.7	68.9	57.5	45.9	33.0	23.2	13.1
	GCE	74.7	75.2	72.7	68.2	64.5	54.9	74.5	72.0	69.6	60.9	50.4	39.5
	JoAPR	64.2	61.5	57.9	53.4	46.8	38.2	63.5	60.2	55.1	49.8	42.1	31.5
	NLPrompt	73.6	74.2	70.5	68.5	65.7	59.6	73.4	72.5	70.5	63.5	57.8	47.4
	OT-AdaGCE	73	72	70.1	69.2	64.3	59.9	73.1	71.2	70.4	64.6	60.1	47
Caltech101	CoOp	81.5	78.4	72.3	61.7	52.2	40.6	81.5	72.6	63.7	47.5	33.1	19.3
	GCE	89.5	89.6	88.7	85.4	81.9	79.2	89.4	88.2	85.6	79.9	69.1	63.9
	JoAPR	79.8	76.4	72.1	68.5	60.2	51.4	78.5	74.1	68.9	62.4	54.7	42.1
	NLPrompt	92.2	91.3	89.9	89.7	88.6	85.6	91.8	91.6	90.5	89.3	85.6	76.1
	OT-AdaGCE	88.6	87.8	87.2	85.1	81.3	80.8	89.4	87.4	86.7	83.9	80.3	71.4

6.2 Evaluation on Real-World Noise

We further evaluated OT-AdaGCE on the Food101N dataset to test its robustness against real-world label noise.

As shown in Table 6, our method achieves an accuracy of **70.9%**. While this outperforms the CoOp baseline (66.9%), it is slightly lower than the reproduced NLPrompt (72.9%). This result aligns with our observation on Caltech101, suggesting that without hyperparameter tuning ($k = 1$), the soft-weighting strategy might be too conservative for certain real-world noise distributions compared to the hard-partitioning strategy.

Table 6: **Real-world Noise Evaluation (Innovation)**

Method	CoOp	GCE	JoAPR	NLPrompt	**OT-AdaGCE**
Accuracy	66.9%	72.7%	72.5%	**72.9%**	70.9%

7 Conclusion

In this project, we successfully reproduced *NLPrompt*, validating its effectiveness against noisy labels in VLMs. From a convex optimization perspective, we rigorously analyzed the OT formulation via Bregman Projections and derived the appropriate method to handle class imbalance. Finally, to overcome the information loss in "hard partitioning", we proposed **OT-Guided Adaptive GCE**, a soft-weighting strategy that dynamically adjusts loss robustness based on optimal transport confidence.

Acknowledgments and Disclosure of Funding

This project was completed for the SI251 Convex Optimization course at ShanghaiTech University, Fall 2025, under the instruction of Prof. Ye Shi.

References

[1] Bikang Pan, Qun Li, Xiaoying Tang, et al. NLPrompt: Noise-Label Prompt Learning for Vision-Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[2] Zhilu Zhang and Mert Sabuncu. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[3] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to Prompt for Vision-Language Models. *International Journal of Computer Vision (IJCV)*, 2022.

[4] Alec Radford, et al. Learning Transferable Visual Models From Natural Language Supervision. In *International Conference on Machine Learning (ICML)*, 2021.

[5] Muhammad Uzair Khattak, Hanoona Rasheed, Muhammad Maaz, Salman Khan, and Fahad Shahbaz Khan. Self-regulating Prompts: Foundational Model Adaptation without Forgetting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[6] Muhammad Uzair Khattak, Hanoona Rasheed, Muhammad Maaz, Salman Khan, and Fahad Shahbaz Khan. MaPLe: Multi-modal Prompt Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[7] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual Prompt Tuning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.