Prior-Aware Multilabel Food Recognition using Graph Convolutional Networks

ILLINOIS

CVPR SEATTLE, WA JUNE 17-21, 2024

META FOOD WORKSHOP

Samyak Rawlekar *, Shubhang Bhatnagar*, Vishnuvardhan Pogunulu Srinivasulu, Narendra Ahuja

Motivation

Tacos Steak

a. Single-Label Data (Food101)

Images from single-label food datasets (as shown in (a)) frequently contain multiple food items. This is also true for real-world food images.

Rice, Asparagus, Chicken Chicken, Beans, Orange

b. Multi-Label Data (FoodSeg103)

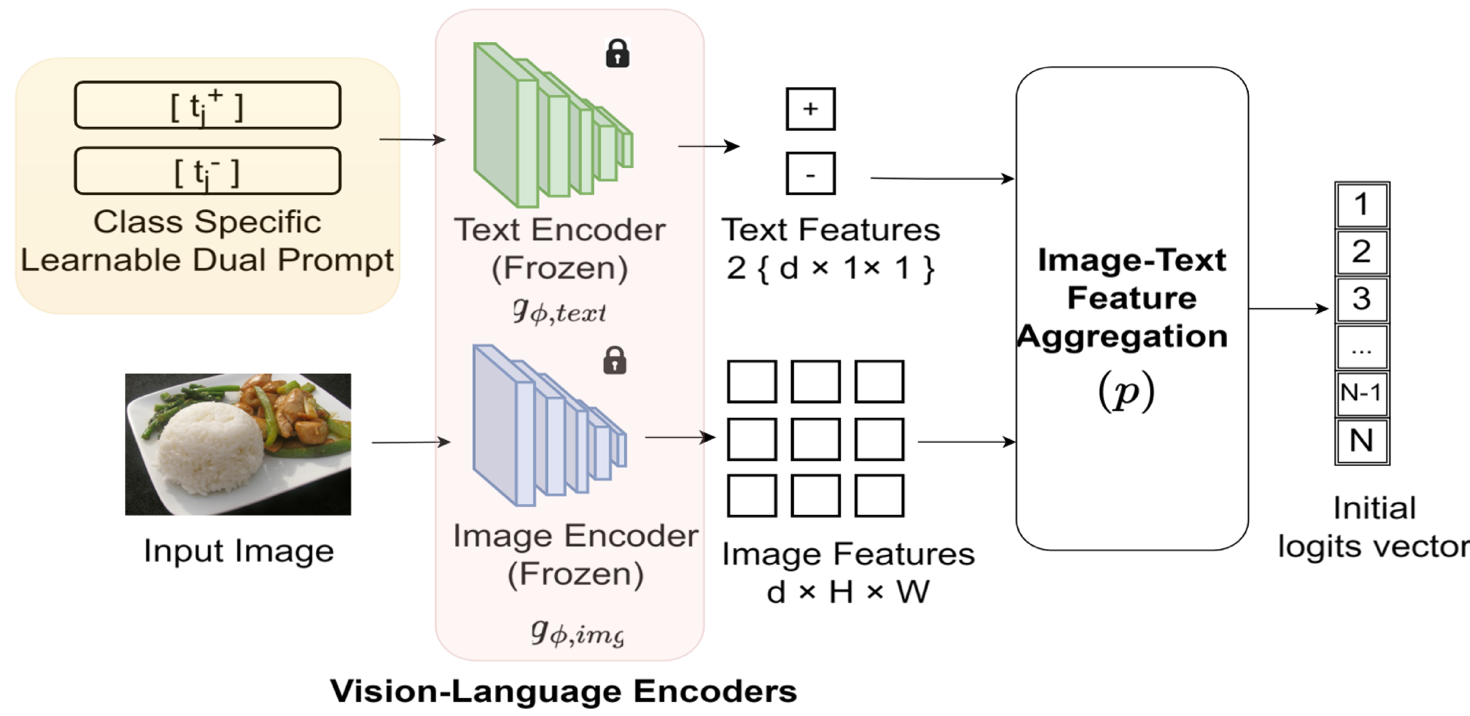
We associate multiple labels with each image (as shown in (b)), casting the problem as Multi-Label Recognition (MLR) where the goal is to identify all items in an image.

Our Contributions

- The occurrences of food items (labels) in dish are correlated. Previous methods detect them independently, thus overlooking valuable co-occurrence information.
- We obtain initial food item logits using CLIP, and refine them to enforce the label correlations seen in training data using a graph convolution network (GCN)
- We propose a new loss function, the Re-weighted Asymmetric Loss (RASL), to address the sample imbalance problem arising from limited food samples in training data.
- Our approach surpasses all SOTA MLR methods on food datasets.

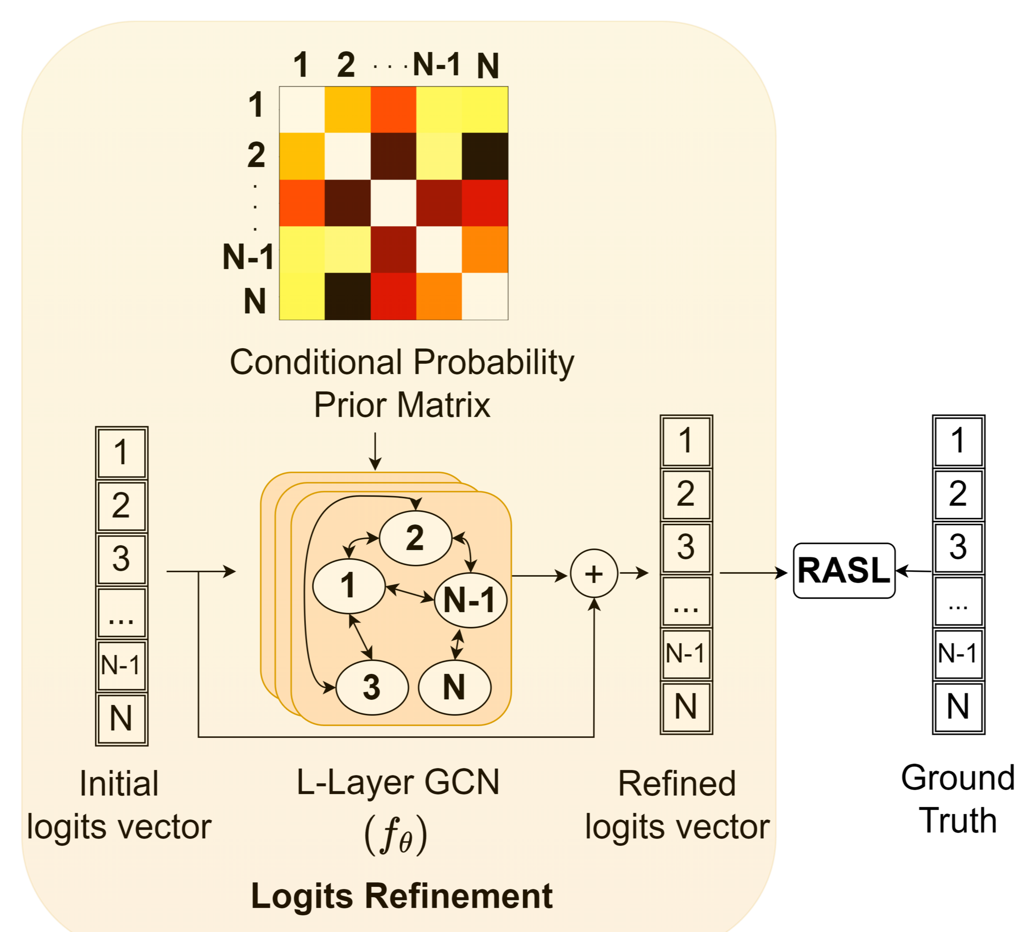
Method

Step -1 : Initial Logits Estimation



- We extract image and text features for all food items by giving the items as prompts to CLIP (g_ϕ).
- Image-text feature aggregation module calculates similarity between text features for each food item and the image features, giving an initial set of logits.

Step -2 : Refining Logits using Conditional Prior



- Initial Logits are refined by a GCN, using the label correlation extracted from the training data.
- We train the framework end-to-end using our proposed Re-weighed ASL (RASL), which helps mitigate class imbalance in the dataset.

Training Loss

We modify the ASL loss by weighting the loss terms of each class by the inverse of the fraction of samples of that class in the total dataset.

Results

Dataset	Method	CP	CR	CF1	mAP
		Metrics (%)			
FoodSeg103 [10]	DualCoOp[9]	43.76	52.54	46.55	48.84
	SCPNet[5]	39.33	54.36	43.14	48.77
	Ours (GCN)	44.76	54.97	47.95	51.24
	Ours (GCN + Reweight)	**48.35**	**55.59**	**49.26**	**52.87**
UNIMIB [4]	DualCoOp[9]	46.37	53.34	48.11	56.01
	SCPNet[5]	50.49	52.85	49.87	59.98
	Ours (GCN)	52.56	59.55	53.78	64.33
	Ours (GCN + Reweight)	**61.39**	**64.62**	**61.42**	**71.19**

Improvement of more than 4% and 11% in mAP on FoodSeg-103 and UNIMIB dataset over SOTA previous MLR works.

How does our method affect performance on classes that are difficult to visually recognize ?

	UNIMIB			FoodSeg103		
	DualCoOp	Ours w/o reweigh	Ours	DualCoOp	Ours w/o reweigh	Ours
CP	25.4	41.9	**57.6**	13.7	14.8	**28.7**
CR	26.2	57.5	**60.0**	19.7	22.5	**26.9**
CF1	24.3	44.9	**59.1**	16.5	18.7	**28.4**

Our approach that models class co-occurrences significantly benefits MLR performance of such classes.
(mean performance of worst 10 classes shown)

Detailed Analysis