ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

MADiMa 2020

S2ML-TL Framework for Multi-Label Food Recognition

Bhalaji Nagarajan[1], Eduardo Aguilar[1,2] and Petia Radeva[1,3]

[1]Universitat de Barcelona, Barcelona, Spain

[2]Universidad Católica del Norte, Antofagasta, Chile

[3]Computer Vision Center, Cerdanyola (Barcelona), Spain

Contents

Introduction

Introduction

- Food Recognition and its challenges
- Transfer Learning
  - Problem of data dependency
  - Negative transfer and Covariate shift
- Multi-Stage TL exploits learned representations from similar easier task on a complex task.

Contributions

- S2ML-TL: Single to Multi-Label Transfer Learning Framework
  - Multi-Label Food Recognition using Single-Label Classifiers

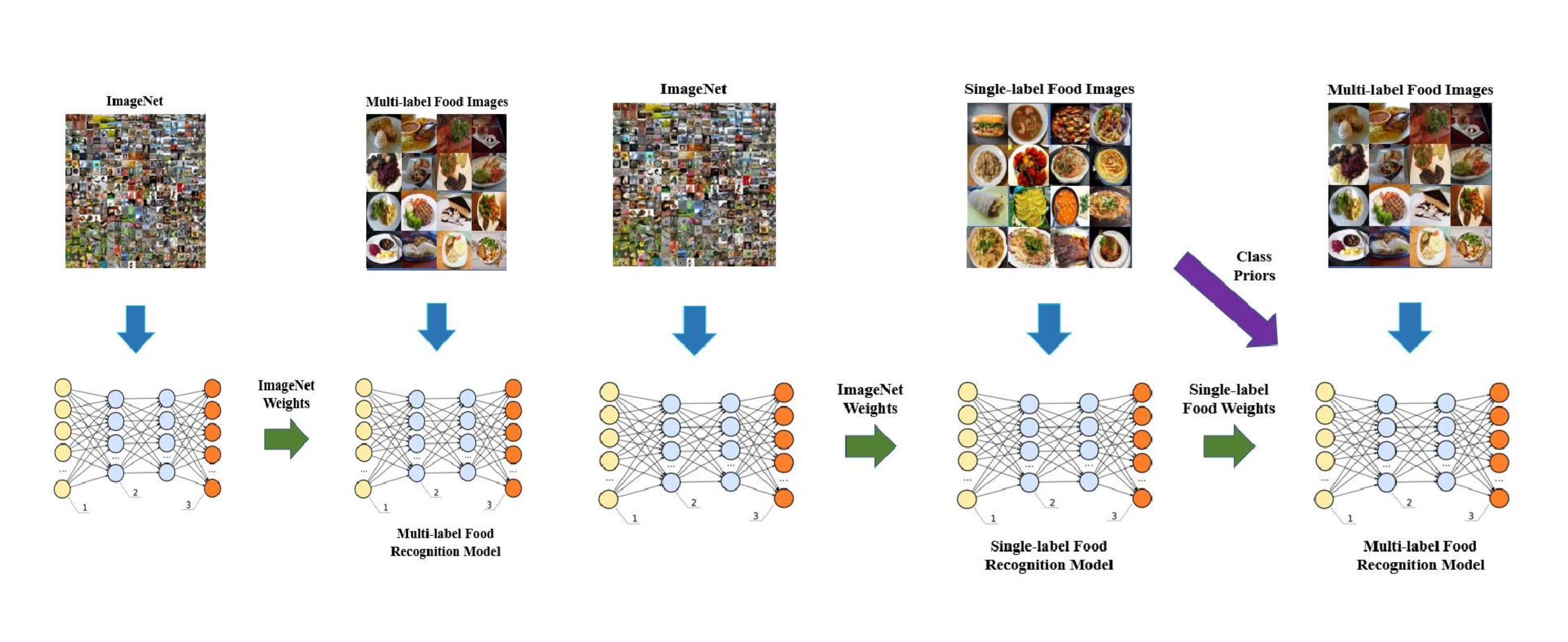
- Class priors to help in selection of positive samples for better transferability

ICPR 2020 25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION Milan, Italy 10 | 15 January 2021

MADiMa 2020

Proposed Methodology

S2ML-TL Framework



Traditional TL (Fine-tuning) vs S2ML-TL Framework

Class Priors

- Generalization of deep networks is difficult due to dataset bias
- Class conditional probabilities define distributions of samples in a domain
- Class priors make learning algorithms more robust to shifts

Prior-Induced Loss Function

- Prior induced loss function *$l_p(x,y)$* is given by

$$l_p(x,y) = \frac{1}{C}\sum_{c=1}^{C}\left[y_i . \log\big(p(y_i)\big) + (1-y) . \log(1-p(y_i))\right] * r_i^b$$

$$r_i^b = \beta\left[\alpha\frac{P_T}{P_S} + (1-\alpha)P_T\right]$$

- P_T/P_S - Ratio of Priors
- α - control the impact of source distribution
- β - correction constant

Prior Computation

$P(T_i)$ - probability of each class, i, in the multi-label datasets

$$P(T_i) = \frac{1}{N_t} \sum_{n=1}^{N_t} y_i^n$$

$P(S_i)$ - sum of probabilities of all classes, where the target class is present

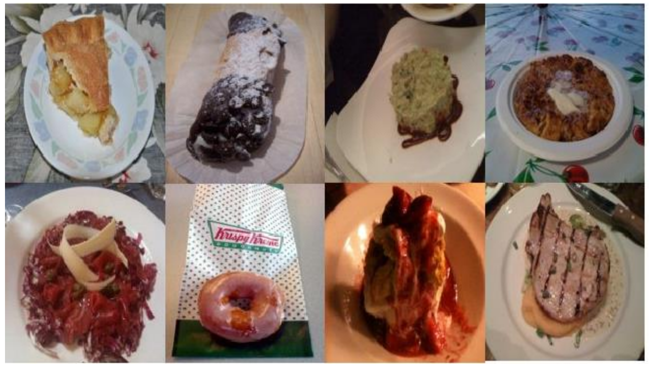
$$P(S_i) = \frac{1}{N_s} \sum_{i} \sum_{j=1}^{N_s} y_j, \qquad j \in \{all\ souce\ classes\ containing\ i\}$$

11

Validation

Datasets and Evaluation Metrics

Single-label **Multi-label**



Food101

Food201

Combo-plates

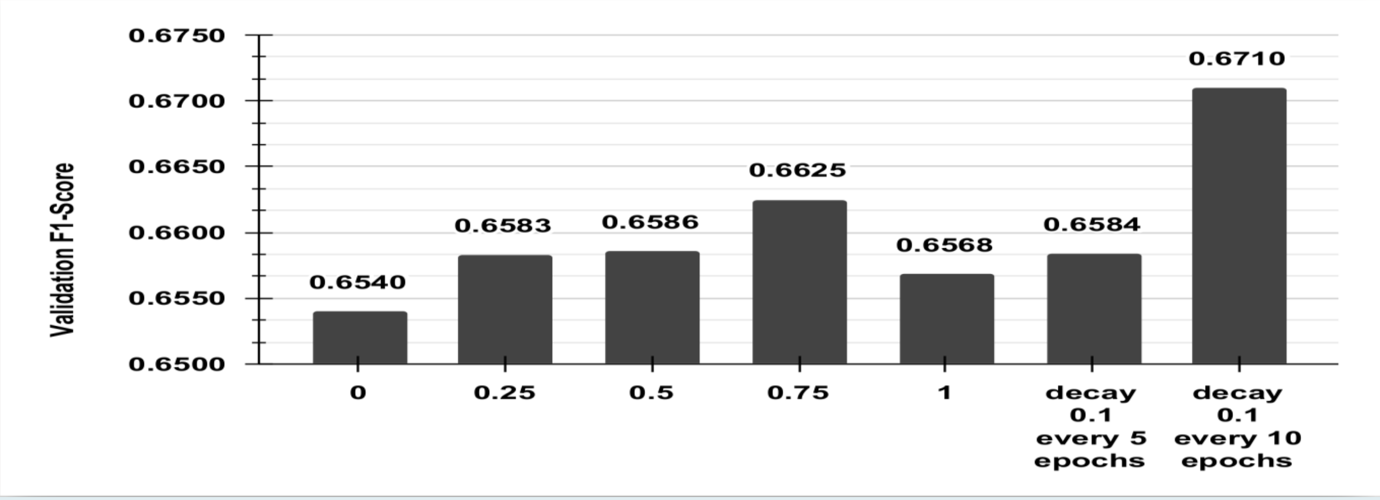
Evaluation Metrics

- Precision, Recall and F1-Score

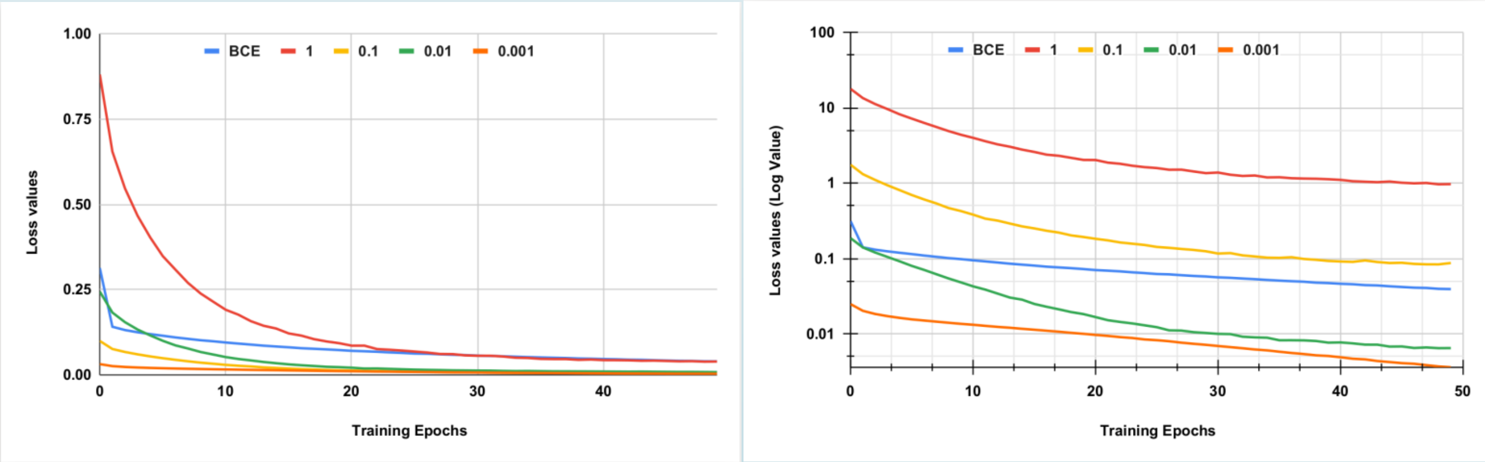
Implementation Details

- Validated using Resnet50 and InceptionResnetV2
- Training single-label images - initialized with ImageNet weights
- Training multi-label images
  - Initialized with ImageNet weights (for standard TL)
  - Initialized with Single-label image weights
    - compared with BCE, prior-induced BCE and KL based BCE
- Parameters
  - Image size: 224 x 224, Batch size: 24
  - Adadelta optimizer with an initial learning rate of 1
- Keras framework with Tensorflow backend

Hyper-parameter Selection



Performance with different α values



Loss curves with varying β values at $\alpha=0$ and $\alpha=0.75$

Results and Discussion

Model Performance of InceptionResnetV2 on Combo-plates

Model	Val. data			Test data		
	Prec.	**Recall**	**F1**	**Prec.**	**Recall**	**F1**
Standard TL	0.7209	0.5865	0.6468	0.7152	0.5840	0.6430
ERM	0.6991	0.5667	0.6260	0.6900	0.5700	0.6200
KL	0.7030	0.6212	0.6596	0.6984	0.6173	0.6553
Priors	0.7045	0.6229	**0.6612**	0.7000	0.6200	**0.6600**

Model Performance of Resnet50 on Combo-plates

Model	Val. data			Test data		
	Prec.	**Recall**	**F1**	**Prec.**	**Recall**	**F1**
Standard TL	0.7250	0.5581	0.6307	0.7200	0.5582	0.6289
BCE	0.7268	0.5590	**0.6320**	0.7223	0.5616	0.6319
KL	0.6956	0.5473	0.6126	0.6933	0.5491	0.6128
Priors	0.6861	0.5783	0.6276	0.6882	0.5886	**0.6345**

Results and Discussion

MADiMa 2020

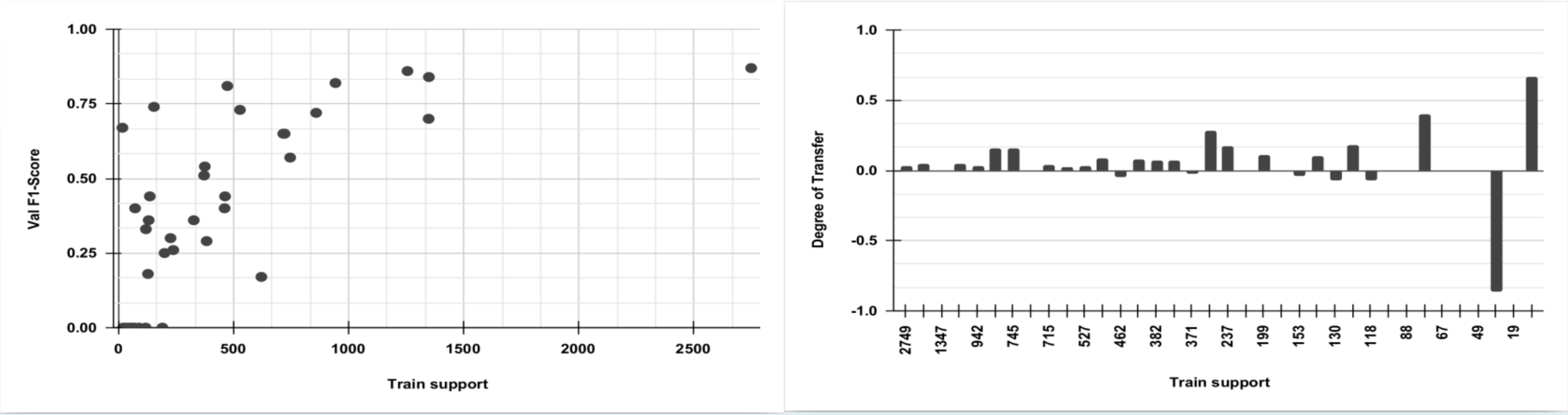
Model Performance of InceptionResnetV2 on Food201

Model	Val. data			Test data		
	Prec.	**Recall**	**F1**	**Prec.**	**Recall**	**F1**
Standard TL	0.6800	0.4595	0.5485	0.6997	0.5001	0.5833
ERM	0.7936	0.5354	0.6394	0.7895	0.5563	**0.6527**
KL	0.7521	0.4755	0.5826	0.7567	0.5044	0.6053
Priors	0.8189	0.6176	**0.7041**	0.7464	0.5550	0.6366

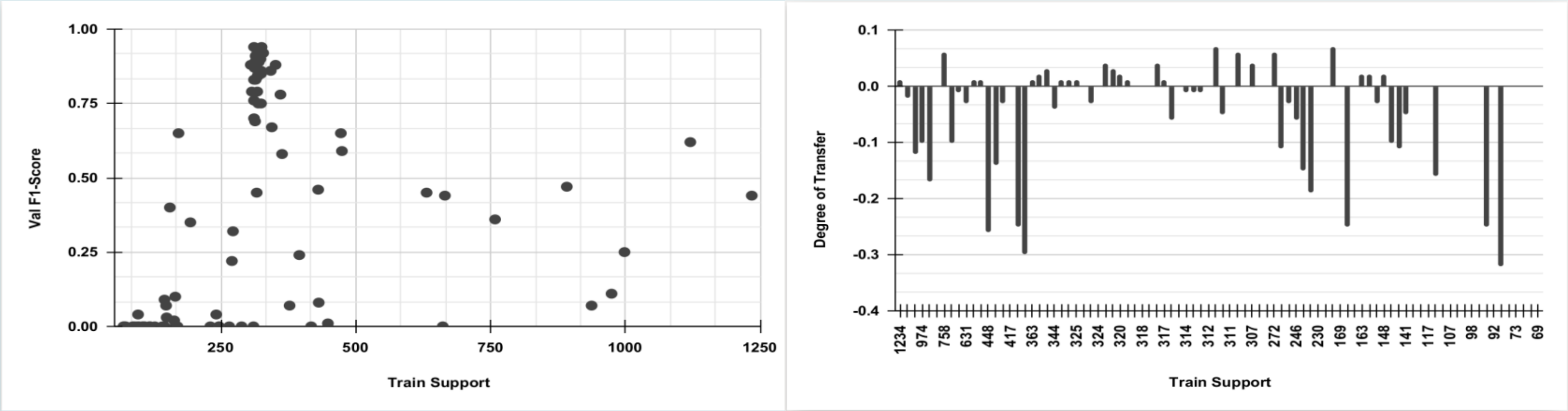
Model Performance of Resnet50 on Food201

Model	Val. data			Test data		
	Prec.	**Recall**	**F1**	**Prec.**	**Recall**	**F1**
Standard TL	0.7204	0.4215	0.5319	0.7322	0.4636	0.5678
ERM	0.7518	0.4546	0.5666	0.7493	0.4800	0.5852
KL	0.7918	0.4317	0.5587	0.7740	0.4370	0.5586
Priors	0.7767	0.5877	**0.6691**	0.7313	0.5400	**0.6213**

Results and Discussion



InceptionResnetV2 on Combo-plates



InceptionResnetV2 on Food201

Ablation Study

Model	a * (P_T/P_S)	P_T	β	Val. data			Test data		
				Prec.	**Recall**	**F1**	**Prec.**	**Recall**	**F1**
a=1 *(w/o β)*	x	-	-	0.7537	0.5685	0.6481	0.7394	0.5666	0.6415
a=1 *(w β)*	x	-	x	0.7237	0.6012	0.6568	0.7200	0.6000	0.6500
Decayed a *(w/o β)*	x	-	-	0.7281	0.5958	0.6553	0.7112	0.5932	0.6469
Decayed a *(w β)*	x	-	x	0.7118	0.6094	0.6567	0.6963	0.6086	0.6495
Target Priors *(w/o β)*	-	x	-	0.6972	0.6238	0.6585	0.6811	0.6140	0.6458
Target Priors *(w β)*	-	x	x	0.7092	0.6068	0.6540	0.7050	0.5994	0.6479
Proposal *(w/o β)*	x	x	-	0.6996	0.6034	0.6480	0.7045	0.6069	0.6521
Proposal *(w β)*	**x**	**x**	**x**	**0.7114**	**0.6349**	**0.6710**	**0.7011**	**0.6127**	**0.6539**

Conclusions

Conclusions

- S2MI-TL framework for multi-label food recognition was achieved using single-label food recognition as an intermediate task
- Validation with two multi-label datasets showed increased learnability of the models
- Class priors further boosted the recognition performance.
- Various hyper-parameter selection decisions were discussed with empyrical evidences.
- Future Scope:
  - Extending the framework for long-tail problem

Acknowledgements

This work was partially funded by TIN2018-095232-B-C21, SGR-2017 1742, Nestore project of the European Commission Horizon 2020 programme (Grant no.769643), Validithi EIT Health, Greenhabit EIT Digital program and CERCAProgramme/Generalitat de Catalunya. We acknowledge the support of NVIDIACorporation with the donation of the Titan Xp GPUs.