





Ontology-Driven Deep Learning Model for Multitask Visual Food Analysis

Daniel Ponte¹, Eduardo Aguilar^{1,2}, Mireia Ribera¹ and Petia Radeva^{1,3}

¹Dept. de Matemàtiques i Informàtica, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, Barcelona, Spain

²Dept. de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, Angamos 0610, Antofagasta, Chile

³Computer Vision Center, Cerdanyola (Barcelona), Spain

Keywords: Food Ontology, Food Image Analysis, Multitask Learning.

Abstract: The food analysis from images is a challenging task that has gained significant attention due to its multiple applications, especially in the field of health and nutrition. Ontology-driven deep learning techniques have shown promising results in improving model performance. Food ontology can leverage domain-specific information to guide model learning and thus substantially enhance the food analysis. In this paper, we propose a new ontology-driven multi-task learning approach for food recognition. To this end, we deal multi-modal information, text and images, in order to extract from the text the food ontology, which represents prior knowledge about the relationship of food concepts at different semantic levels (e.g. food groups and food names), and apply this information to guide the learning of the multi-task model to perform the task at hand. The proposed method was validated on the public food dataset named MAFood-121, specifically on dishes belonging to Mexican cuisine, outperforming the results obtained in single-label food recognition and multi-label food group recognition. Moreover, the proposed integration of the ontology into the deep learning framework allows providing more consistent results across the tasks.

1 INTRODUCTION


Food recognition from images has gained significant attention due to its various applications, including diet tracking (Ming et al., 2018), food recommendation (Deldjoo et al., 2020), and health analysis (Allegra et al., 2020). Despite this, food recognition remains a challenging task due to the complexity of food images, which can vary in terms of their appearance, size, shape, texture, and color (Jiang et al., 2019). Additionally, food may be presented in different settings, such as plates, bowls, or trays, adding another layer of complexity to the classification task. Therefore, a well-complied deep learning model are required to correctly address the classification of food images.


In recent years, the use of ontologies in the classification of food images has shown promising results (Zhao et al., 2021; Wang et al., 2022), which has led to the development of ontology-based image classification


techniques. However, the lack of a standardized food ontology is a major problem to food between different systems and applications. This can lead to ambiguity and confusion in how food is classified and described, making it difficult to integrate with different systems. Additionally, it can have a negative impact in areas such as public health, nutrition research, agriculture, and the food industry, among others, where accurate and complete information about food is important.


The integration of the food ontology into a deep learning framework could provide several benefits such as: a) Relating the food concepts predicted by the model to specific diseases to prevent damage to the health of people suffering from it (Donadello and Dragoni, 2019); b) Relating the food concepts predicted by the model with allergens to prevent food allergy or intolerance; c) Training of multi-task model being aware of the coexistence and exclusion of food concepts at different semantic levels (Wang et al., 2022), just to mention a few.

Ontology-based deep learning models (Popovski et al., 2020) have been proposed that provide an overview and comparison of named entity recog-

^a <https://orcid.org/0000-0002-4482-3645>

^b <https://orcid.org/0000-0002-2463-0301>

^c <https://orcid.org/0000-0003-1455-1869>

^d <https://orcid.org/0000-0003-0047-5172>

nition methods in the food domain, which can be used for automated information extraction about food from text, where four methods are discussed: FoodIE (Popovski et al., 2019), NCBO (SNOMED CT), NCBO (OntoFood) and NCBO (FoodON) (Stojanov et al., 2020). The comparison is performed using a dataset of 1000 recipes taken from Allrecipes (Song et al., 2023). They have also been proposed on extracting food information from text (Popovski et al., 2020) or substituting ingredients in recipes (Ławrynowicz et al., 2022). Although these approaches are interesting, they do not directly focus on classifying food images.

In the area of large-scale visual recognition, several approaches have been proposed that combine ontologies and deep learning (Kuang et al., 2018; Zhang et al., 2019). In particular, in the field of food image recognition, lately, there has been a growing interest. In (Divakar et al., 2019), the problem of predicting Type 2 Diabetes Mellitus is addressed by proposing an ontology-based model to improve the accuracy of food recognition algorithms. In (Wang et al., 2022) proposes an ingredient ontology and joint learning, in (Zhao et al., 2021) presents a fusion learning framework with semantic embedding, in (Donadello and Dragoni, 2019) focuses on ontologies and deep neural networks, and in (Kuang et al., 2018) introduces multi-level deep learning.

Our method differs from previous approaches in that it focuses on building a well-structured food ontology to enable knowledge transfer from it to a neural network, with the aim of leveraging multi-modal data (text and image) for uni-modal multitask recognition of food images. The idea behind this work is that an adequate ontology is essential for the accurate classification of foods. By building a strong ontology, a clear and coherent structure can be established to represent knowledge about food and its properties. This ontology can be used to guide the classification of foods through a neural network, allowing greater precision and a better understanding of the characteristics of each food. Moreover, transferring knowledge from the ontology to the neural network can also help improve the generalization capability of the network.

Our main contributions are:

- Careful building of a specialized food ontology from textual data retrieval in public food recipes.
- Ensure consistency in multi-tasking results by integrating prior knowledge extracted from the ontology into a food image classification model.
- The proposed ontology-driven method improved performance at both the dish and food group level compared to the baseline approach.

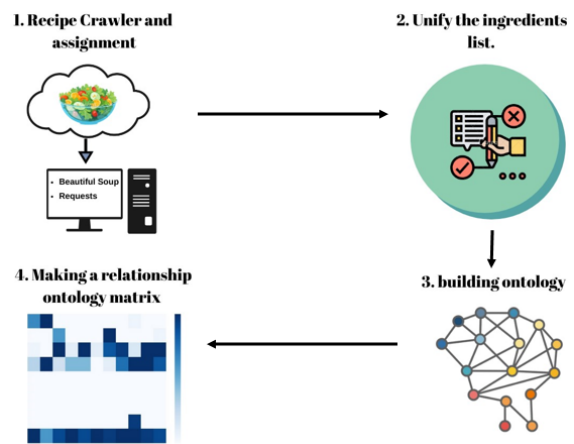


Figure 1: Ontology creation diagram.

2 METHODOLOGY

This section describes the process of building a food ontology and its application in the classification of food groups and dishes.

2.1 Food Ontology

Food ontology is governed as a fundamental pillar of this methodology. In the building of the proposed food ontology, we focus on two semantic levels of food concepts: the dish name (e.g., caesar salad, guacamole) and the food groups (e.g., vegetables, bread). An illustration of the stages involved in the proposed process can be seen in Fig. 1. First, information is collected from various web sources, including food websites. From these resources, lists of recipes covering a wide range of food dishes are obtained. To carry out the construction of the ontology, the most representative recipes for each dish are extracted. The ontological hierarchy is established by grouping ingredients at different semantic levels that allow classifying and relating food groups to dishes. This process involves converting ingredients into more general categories, following the guidance provided by the Heliis food ontology (Dragoni et al., 2018). This step significantly simplifies the organization and search of the food data. The resulting ontology is nourished with precise and coherent relationships between food groups and dishes.

Each stage is described in more detail in the following subsections.

2.1.1 Recipe Crawler and Assignment

The first stage consists of collecting recipes related to a specific dish from food web pages enriched with

google recipes metadata such as Yummly (Yummly, 2023) and AllRecipes (Allrecipes, 2023), to identify the ingredients commonly used in its preparation. The retrieval is performed with Beautiful Soup Python library using web scraping strategies and taking benefit of the structured information. From that, a list of R recipes is compiled for each food dish taking into account the similarity between the dish name and the recipe title.

2.1.2 Unification of the Ingredients List

Once all the recipes are linked to the dishes belonging to the target dataset, the next step is to extract a list of the ingredients from the information provided in the recipes. For this purpose, natural language processing techniques (available in the NLTK library) are applied to analyze the text, including the removal of anything that differs from the ingredient names as part of the normalization process (e.g. stopwords, verbs, culinary measurements, etc.). As a result, a unique list of ingredients is created. This list is generated to facilitate the construction of the ontology and to normalize the data, which is essential for the subsequent stage. Finally, the unique list of ingredients is subjected to a further refinement process. Selective removals are manually made on ingredients that are not visible or would not be used in the experiments (e.g., salt, pepper, Vinegar). This process is essential to ensure that the ontology is composed of only relevant ingredients.

2.1.3 Building Ontology

In this stage, the recipes and their corresponding ingredients are linked to each dish. In fact, the ingredients are grouped into a high-level food concept (food groups) such as "meat," "vegetables," "fruits," "cereals," "dairy products," "spices," among others identified through SPARQL queries on the Helis ontology (Donadello and Dragoni, 2019) with the GraphDB tool (Güting, 1994). These food groups may be different from those available in the target dataset. Therefore, an additional step linking the food group to the available annotations is necessary. This step is performed manually in our experiments where for example food groups such as 'beans' and 'fruits' were considered vegetables to preserve the original annotations of the target dataset.

2.1.4 Making a Relationship Ontology Matrix

To deepen the analysis and exploitation of the ontology, a coexistence matrix is created. This matrix captures the relationship between food groups and dishes. However, it is highlighted that this structure can be

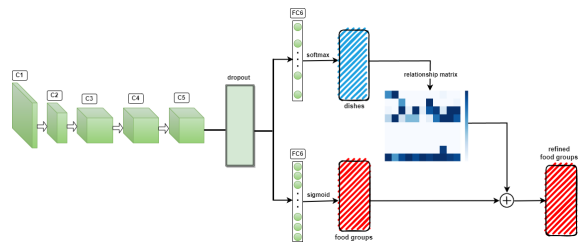


Figure 2: The framework of the proposed method.

scaled to include more semantic levels, which would allow a more detailed and granular representation. The central purpose of the matrix is to quantify the presence of food groups in each of the dishes. Each cell in the matrix stores the number of times a specific food group appears in a particular dish. This provides a quantitative view of the composition of each dish in terms of food groups.

The relationship between semantic levels is reflected in the structure of the relationship matrix (see equation 1). The columns represent the names of the dishes, while the rows refer to the food groups. The values in the cells of the matrix indicate the strength of the relationship between food groups and dishes. This reflects not only how many times a food group is found in a dish, but also the diversity of food groups present on it. The matrix also allows to understand the relationship between the semantic levels of the food groups and their distribution in the dishes.

The relationship matrix RM is formally defined as follows:

$$RM = \begin{bmatrix} \frac{\sum_{r=1}^{|R^1|} |FG_1 \in R_r^1|}{|R^1|} & \dots & \frac{\sum_{r=1}^{|R^D|} |FG_1 \in R_r^D|}{|R^D|} \\ \dots & \frac{\sum_{r=1}^{|R^d|} |FG_g \in R_r^d|}{|R^d|} & \dots \\ \frac{\sum_{r=1}^{|R^1|} |FG_G \in R_r^1|}{|R^1|} & \dots & \frac{\sum_{r=1}^{|R^D|} |FG_G \in R_r^D|}{|R^D|} \end{bmatrix} \quad (1)$$

where $|R^d|$ corresponds to the number of recipes linked to the d -th dish, R_r^d - the list of food groups for the r -th recipe linked to the d -th dish and FG_g - the g -th food group.

2.2 Ontology-Driven Multitask Food Recognition

The proposed ontology-driven deep learning method for performing multi-task food recognition is illustrated in Fig. 2. In deep learning, a multi-task approach can be performed from a generic network, where all parameters are shared to extract features, to a specific network, where independent networks are used for each task (Misra et al., 2016). For multi-task

food recognition, a generic network is considered due to the similarity of the task at hand (food recognition and food groups recognition), where both can benefit from the extracted general features from the backbone (e.g., ResNet50 (He et al., 2016)). At the top of the backbone, a dropout layer is considered to avoid overfitting, followed by a specific fully connected layer for each task.

For food recognition, softmax activation is applied on the logits layer to provide a probability of the most likely dish. For food group recognition, like any multi-label task, a sigmoid activation is applied on the logits layer to provide independent probability for each group. Afterwards, the probability of each food concept belonging to each task, is obtained. An essential component of this network is the integration of a food ontology, specifically the relationship matrix, which acts as an additional layer that reflects the hierarchical and semantic relationships between food groups and dishes. The output of this layer is aggregated to the output of the food groups to provide a refined probability of them. This ontology provides contextual information to the network, allowing a better understanding of the composition of the dishes and a more precise classification. The interaction between the ontology and the network is achieved by a custom layer, which weights the model predictions based on the relationships established in the ontology.

The probability of the refined food groups is formally defined as follows:

$$p(y_g|W, RM) = \lambda \cdot p(y_g|W) + (1 - \lambda) \cdot \sum_{d=1}^D p(y_d = d|W) \cdot p(y_g|y_d = d), \quad (2)$$

$$p(y_g|W) = \frac{1}{1 + e^{f_g^W(x)}}, \quad (3)$$

$$p(y_d|W) = \frac{e^{f_d^W(x)_d}}{\sum_{k=1}^K e^{f_d^W(x)_k}}, \quad (4)$$

$$p(y_g|y_d) = RM[g, d], \quad (5)$$

where $p(y_g|W)$ represents the conditional probability that a specific ingredient y_g is present in the image; W - the model weights; $f_g^W(x)$ - the logits outputs for the food groups; $p(y_d|W)$ - the probability that a specific dish y_d is the correct class; $f_d^W(x)_d$ - the d-th logits output for the dishes; $p(y_g|y_d)$ - the probability, extracted from the relationship matrix of having a food group y_g given the dish y_d ; K and D - the number of dishes; x - the input image; and λ - a hyperparameter to weight the contribution of the both terms.

For the model learning, two equally weighted loss functions are used: Cross-Entropy Loss (CELoss) for dish task and Binary Cross-Entropy Loss (BCELoss) for the food groups task. The equation representing CELoss function for a single input image is expressed as follows:

$$CELoss = - \sum_{d=1}^D \hat{y}_d \cdot \log(p(y_d|W)), \quad (6)$$

where \hat{y}_d is the Ground Truth (GT) label in one-hot encoding and $p(y_d|W)$ is the probability given by the model for the d-th dish. The CEloss is calculated individually for each image and then averaged.

Regarding the BCELoss, it is formally defined as follows:

$$BCELoss = \sum_{g=1}^G \hat{y}_g \cdot \log(p(y_g|W, RM)) + \sum_{g=1}^G (1 - \hat{y}_g) \cdot \log(1 - p(y_g|W, RM)) \quad (7)$$

where G represents the number of food groups; \hat{y}_g is the GT in one-hot encoding for g-th food group, indicating whether food group g is present or not in the sample; and $p(y_g|W, RM)$ is the refined probability given for g-th food group. The BCEloss is calculated individually for each image and then averaged over the total images and food groups.

It is interesting to note that although the proposed method explicitly refines the prediction of food groups, the fact of using the probability of the dish, together with the relationship matrix, for the refinement also indirectly results in the predictions for the dishes being refined.

3 VALIDATION

In this section, we present the dataset, the experimental setting, and the different evaluation metrics.

3.1 Dataset

The data set used in this research, known as MAFood-121 (Aguilar et al., 2019), consists of a total of 21,175 images representing traditional dishes from eleven of the most popular cuisines in the world. These dishes, which reflect the richness and diversity of global gastronomy, have been grouped into a dataset covering 121 dishes in total. Each of the images belonging to these dishes is labeled with at least one of the ten food groups previously defined: bread, egg, fried foods, meat, noodles/pasta, rice, seafood, soup, dumpling and vegetables.

For the purposes of this research, it is decided to focus on Mexican cuisine, selecting 11 representative dishes with a total of 2,242 images that encapsulate the essence and culinary variety of this tradition. These dishes range from classics like 'caesar salad', 'enchiladas' and 'tacos', to delicacies like 'guacamole', 'pozole' and 'tostadas'. This meticulous selection is carried out with the purpose of focusing the classification task in a specific and representative context, which allows a detailed and precise analysis of the Mexican gastronomic wealth. In the experiments, we maintain the original division of the dataset. Specifically, 73.6% of the images are used for training, 12.44% for validation and the remaining 13.96% for testing.

3.2 Experimental Setup

For the implementation of the proposed method, ResNet50 is selected as the backbone. This network is also used as a baseline for comparison purposes. ResNet50 is pretrained on ImageNet and then the baseline and proposed method are retrained for a total of 20 epochs, using a set of empirically selected hyperparameters. A learning rate (LR) of 0.001 is used and the batch size is set to 64 to balance computational efficiency and training stability. As part of the regularization, a Dropout layer after the last convolutional layer with a rate of 0.1 is included to avoid overfitting in both methods. The inclusion of this layer provides effective regulation and improves the model's ability to generalize to unseen data. The Adam optimizer is used to minimize the loss function. Additionally, a simple data preprocessing is applied that included image resizing to 224x224 pixels and normalization with a mean and standard deviation of 0.5. On the other hand, the number of recipes for each dish is set at 20 ($|R^d| = 20$), because we detect that as we increase this number, recipes are recovered that do not represent well the dish consulted. Finally, λ is set from 0.5 to 0.9 with a step of 0.1.

Traditional evaluation metrics were selected for each of the target tasks. For the single-label food recognition problem, whose aim is to classify the general context of the images with the most likely food, the Accuracy metric was used. For the multi-label classification problem of food group recognition, that involves categorizing food images with food groups representing each ingredient contained in the food, four metrics was selected: 1) Precision (P) which corresponds the proportion of correctly predicted instances relative to the total number of instances predicted under that label, 2) Recall (R) which measures the model's ability to capture all true labels, 3) F1

score which combines precision and recall into a single metric, providing a comprehensive assessment of model performance and 4) The Jaccard index which evaluates the overlap between true and predicted labels, quantifying the degree of similarity in the set of labels. Furthermore, the metric Multi-task accuracy (MTA) (Aguilar et al., 2019) was used to quantify the consistency of model predictions across multiple classification tasks.

4 RESULTS

In this work, we evaluate the performance of the multi-task classification methods applied to the challenge of identifying food dishes and their food groups from images. For this purpose, ResNet50 and the proposed ontology-driven ResNet50 (OD-ResNet50) were evaluated on Mexican dishes belonging to the MAFood-121 dataset. Regarding OD-ResNet-50, the parameter λ is set after analysis of the results on the validation set. Five experiments were performed changing λ from 0.5 to 0.9. The results obtained for the training and validation set can be seen in Fig. 3. As observed, low λ tends to provide lower performance than high λ . We identify that a λ of 0.8 or 0.9 are appropriate and are therefore used to evaluate the performance of the model on the test set.

Table 1 summarizes the results obtained by evaluating the performance of three models: ResNet50, OD-ResNet50 with λ equals to 0.9 and OD-ResNet50 with λ equals to 0.8. The evaluations were performed on the Test set, Validation set and Training set. The results are expressed in terms of Precision (P), Recall (R), F₁-Score (F_1), Jaccard Index, Accuracy and the MTA. In the Validation and Test sets, it is observed that the proposed method outperforms the baseline in all the metrics evaluated in both food recognition and food groups recognition and also in the joint evaluation (MTA). On the other hand, it can also be observed that when we compare the proposed method using λ equal to 0.8 with respect to λ equal to 0.9, we observe that a low λ provides a better P for the food groups while a high λ provides better results in all remaining metrics. This suggests that the more strict the ontological integration, the model tends to provide fewer false predictions although it loses its ability to find all food groups. As for the Training set, a different behavior is observed. In this case, the performance of the food groups is slightly higher than the proposed method, although the accuracy for food recognition is still lower. These results demonstrate the ability of the proposed method to generalize better, particularly with respect to food groups.

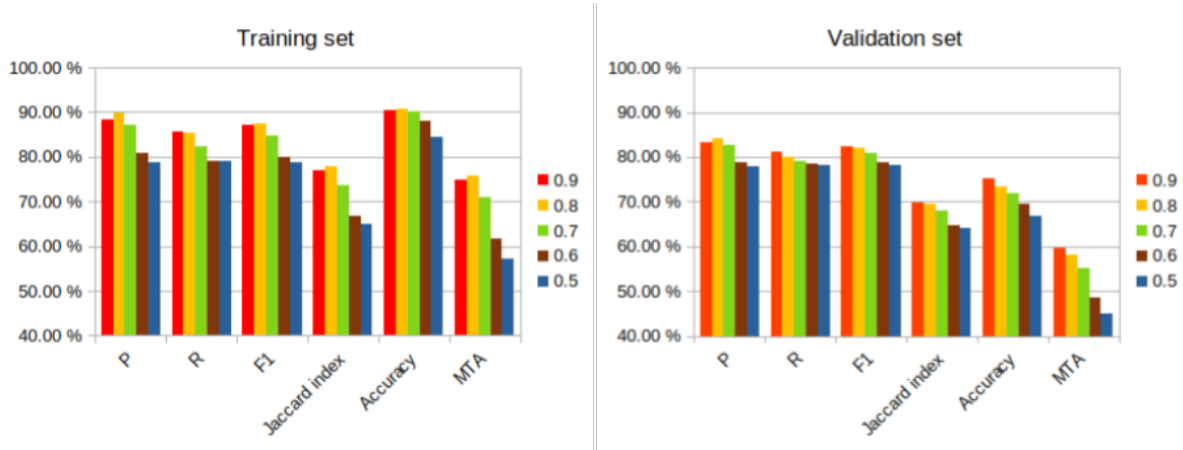


Figure 3: Performance of OD-ResNet50 on the training and validation sets using λ from 0.5 to 0.9.

Table 1: Comparison of the performance of classification methods with single-label, multi-label and multitask metrics.

Method	λ	P	R	F ₁	Jaccard index	Accuracy	MTA
Test set							
ResNet50	-	0.8439	0.8311	0.8374	0.7203	0.6741	0.5271
OD-ResNet50	0.9	0.8498	0.8356	0.8427	0.7281	0.7093	0.5536
OD-ResNet50	0.8	0.8562	0.8250	0.8403	0.7246	0.6997	0.5531
Validation set							
ResNet50	-	0.8289	0.8034	0.8160	0.6891	0.7419	0.5820
OD-ResNet50	0.9	0.8333	0.8120	0.8225	0.6985	0.7527	0.5972
OD-ResNet50	0.8	0.8417	0.8000	0.8203	0.6954	0.7348	0.5826
Training set							
ResNet50	-	0.8887	0.8685	0.8785	0.7833	0.8916	0.7477
OD-ResNet50	0.9	0.8847	0.8564	0.8703	0.7704	0.9055	0.7472
OD-ResNet50	0.8	0.8983	0.8539	0.8755	0.7786	0.9073	0.7570

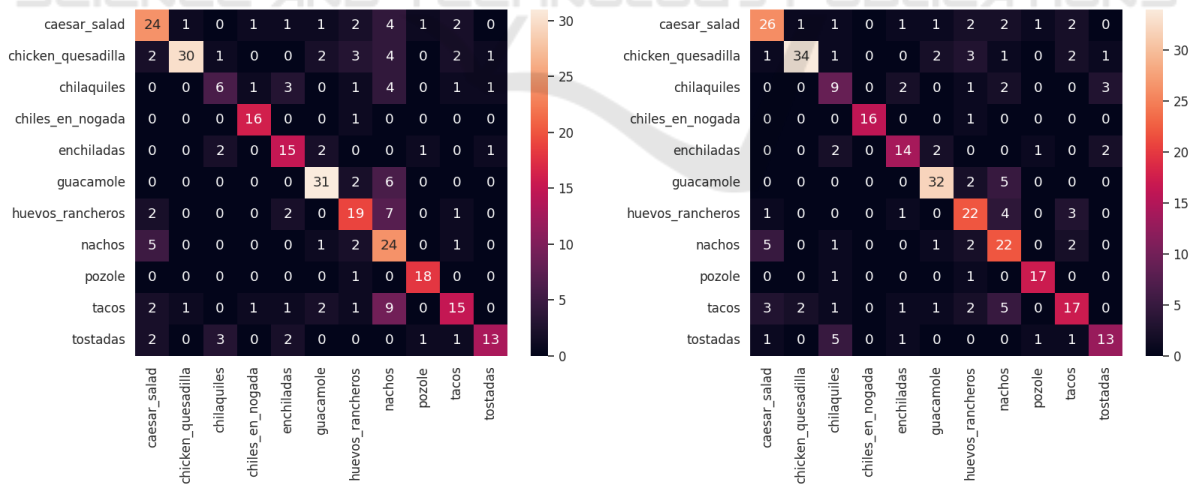


Figure 4: Confusion matrix of the food recognition provided by ResNet50 (left) and OD-ResNet50 (right) with λ equals to 0.9.

Figure 4 shows the confusion matrix related to food recognition provided by the baseline and the proposed method. In general, it is observed that in most dishes the OD-ResNet50 model provides equal or better performance than ResNet50. We can also see that

OD-ResNet50 is less sensitive to class imbalance. In particular, this can be noticed in the chilaquiles plate, which contains the smallest number of images in the data set. In this case, the model was able to misclassify the images much less than ResNet50.





	GT	OD-ResNet50	ResNet50
	Dish: enchiladas Food groups: bread	Dish: enchiladas 0.81 Food groups: bread 0.90 vegetable 0.64	Dish: enchiladas 0.80 Food groups: bread 0.99 meat 0.55 vegetable 0.68
	Dish: huevos_rancheros Food groups: egg, vegetable	Dish: huevos_rancheros 0.80 Food groups: egg 0.65 vegetable 0.72	Dish: huevos_rancheros 0.71 Food groups: bread 0.60 egg 0.61 vegetable 0.96
	Dish: pozole Food groups: soup, vegetable	Dish: pozole 0.76 Food groups: soup 0.56 vegetable 0.89	Dish: pozole 0.72 Food Groups: vegetable 0.87
	Dish: guacamole Food groups: bread, vegetable	Dish: guacamole 0.98 Food groups: bread 0.47 vegetable 1.00	Dish: guacamole 0.86 Food groups: bread 0.62 vegetable 0.99

Figure 5: Success and failure cases of OD-ResNet50 with λ equals to 0.9 on MAFood-121.

Qualitative results of ResNet50 and OD-ResNet50 model are presented in the Fig. 5. From this, the positive influence of the ontology on the performance of the classification models can be clearly observed. In the results obtained by OD-ResNet50 for the first three examples a significant increase in the accuracy of the predictions for both food dishes and food groups is observed. For example, in the case of 'enchiladas', the model with ontology achieves a confidence score of 0.81, while the ResNet50 model (without ontology) achieves a score of 0.80, indicating a substantial improvement in prediction ability. This pattern is repeated in other examples, such as 'huevos rancheros' and 'pozole', where the OD-ResNet50 model clearly outperforms the ResNet50 model. This reinforces the idea that the inclusion of an ontology provides additional and consistent information that supports the classification process, thus improving the accuracy of the predictions. However, in few cases we noticed a negative influence of the ontology. For example, in the case of 'guacamole', the ResNet50 model slightly outperforms the model with ontology in food group classification. The reason for this is that the food group bread is not common for guacamole and for this reason the model must be very secure so as not to lose that prediction due to prior knowledge incorporated by the ontology.

In summary, the inclusion of an ontology in the classification process brings substantial improvements in most cases, suggesting its relevance in future

applications in the field of computer vision and food classification.

5 CONCLUSIONS

This work highlights the relevance of ontologies in food classification from images. Incorporating the ontology into the deep learning model resulted in a substantial improvement in accuracy in the classification of food groups and food dishes. Additionally, greater consistency in responses and more effective knowledge transfer from the ontology to the learning model were observed from the results, improving its generalization ability. Furthermore, the proposed multimodal information integration, which combines text and image data, enriches the model learning process. This combination allowed the model to acquire knowledge more accurately and effectively, resulting in greater accuracy in food classification. In terms of future research, a wide spectrum of possibilities opens up. Integration of additional information, such as nutritional data, could be explored in order to further improve accuracy in food classification. Likewise, it could be investigated how the ontology could be used in the detection of allergens in food, which would have fundamental implications for food safety. Additionally, the applications of ontology and multi-modal information in food classification in various cultures and regions of the world could be studied.

ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish project PID2022-136436NB-I00 (AEI-MICINN), Horizon EU project MUSAE (No. 01070421), 2021-SGR-01094 (AGAUR), Icrea Academia'2022 (Generalitat de Catalunya), Robo STEAM (2022-1-BG01-KA220-VET-000089434, Erasmus+ EU), DeepSense (ACE053/22/000029, ACCIÓ), DeepFoodVol (AEI-MICINN, PDC2022-133642-I00), PID2022-141566NB-I00 (AEI-MICINN), CERCA Programme / Generalitat de Catalunya, and Agencia Nacional de Investigación y Desarrollo de Chile (ANID) (Grant No. FONDECYT INICIACIÓN 11230262). D. Ponte acknowledges the support of Secretaría Nacional de Ciencia, Tecnología e Innovación Senacyt Panamá (Scholarship No. 270-2022-125).

REFERENCES

- Aguilar, E., Bolaños, M., and Radeva, P. (2019). Regularized uncertainty-based multi-task learning model for food analysis. *Journal of Visual Communication and Image Representation*, 60:360–370.
- Allegra, D., Battiato, S., Ortis, A., Urso, S., and Polosa, R. (2020). A review on food recognition technology for health applications. *Health Psychology Research*, 8(3).
- Allrecipes, I. (2023). Allrecipes.
- Deldjoo, Y., Schedl, M., Cremonesi, P., and Pasi, G. (2020). Recommender systems leveraging multimedia content. *ACM Computing Surveys (CSUR)*, 53(5):1–38.
- Divakar, H., Ramesh, D., and Prakash, B. (2019). An ontology driven system to predict diabetes with machine learning techniques. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9:4005–4011.
- Donadello, I. and Dragoni, M. (2019). Ontology-driven food category classification in images. In *ICIAP*, pages 607–617. Springer.
- Dragoni, M., Bailoni, T., Maimone, R., and Eccher, C. (2018). Helis: An ontology for supporting healthy lifestyles. In *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part II 17*, pages 53–69. Springer.
- Güting, R. H. (1994). Graphdb: Modeling and querying graphs in databases. In *VLDB*, volume 94, pages 12–15. Citeseer.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *CVPR*, pages 770–778.
- Jiang, S., Min, W., Liu, L., and Luo, Z. (2019). Multi-scale multi-view deep feature aggregation for food recognition. *IEEE Transactions on Image Processing*, 29:265–276.
- Kuang, Z., Yu, J., Li, Z., Zhang, B., and Fan, J. (2018). Integrating multi-level deep learning and concept ontology for large-scale visual recognition. *Pattern Recognition*, 78:198–214.
- Ławrynowicz, A., Wróblewska, A., Adrian, W. T., Kulczyński, B., and Gramza-Michałowska, A. (2022). Food recipe ingredient substitution ontology design pattern. *Sensors*, 22(3):1095.
- Ming, Z.-Y., Chen, J., Cao, Y., Forde, C., Ngo, C.-W., and Chua, T. S. (2018). Food photo recognition for dietary tracking: System and experiment. In *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part II 24*, pages 129–141. Springer.
- Misra, I., Shrivastava, A., Gupta, A., and Hebert, M. (2016). Cross-stitch networks for multi-task learning. In *CVPR*, pages 3994–4003.
- Popovski, G., Kochev, S., Korousic-Seljak, B., and Eftimov, T. (2019). Foodie: A rule-based named-entity recognition method for food information extraction. In *ICPRAM*, pages 915–922. SCITEPRESS.
- Popovski, G., Seljak, B. K., and Eftimov, T. (2020). A survey of named-entity recognition methods for food information extraction. *IEEE Access*, 8:31586–31594.
- Song, Y., Yang, X., and Xu, C. (2023). Self-supervised calorie-aware heterogeneous graph networks for food recommendation. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(1s):1–23.
- Stojanov, R., Kocev, I., Gramatikov, S., Popovski, G., Seljak, B. K., and Eftimov, T. (2020). Toward robust food ontology mapping. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 3596–3601. IEEE.
- Wang, Z., Min, W., Li, Z., Kang, L., Wei, X., Wei, X., and Jiang, S. (2022). Ingredient-guided region discovery and relationship modeling for food category-ingredient prediction. *IEEE Transactions on Image Processing*, 31:5214–5226.
- Yummly, I. (2023). Yummly.
- Zhang, Y., Qu, Y., Li, C., Lei, Y., and Fan, J. (2019). Ontology-driven hierarchical sparse coding for large-scale image classification. *Neurocomputing*, 360:209–219.
- Zhao, H., Yap, K.-H., and Kot, A. C. (2021). Fusion learning using semantics and graph convolutional network for visual food recognition. In *WACV*, pages 1711–1720.