International Workshop on Embedded Artificial Intelligence Devices, Systems, and Industrial Applications (EAI)

Milan, Italy 19 September 2022

International Workshop on Embedded Artificial Intelligence Devices, Systems, and Industrial Applications (EAI)

ANDANTE

AI ECSEL JU

AI4DI

TEMPO

Food Ingredients Recognition Through Multi-label Learning

MILAN
19-22 SEPTEMBER 2022
ESSDERC
ESSCIRC

19 September 2022 Milan, Italy

Presentation Outline

- Introduction
- Background
- Research Goal — Focus
- Design and Implementation
- Experimental Results
- Conclusions & Future Work

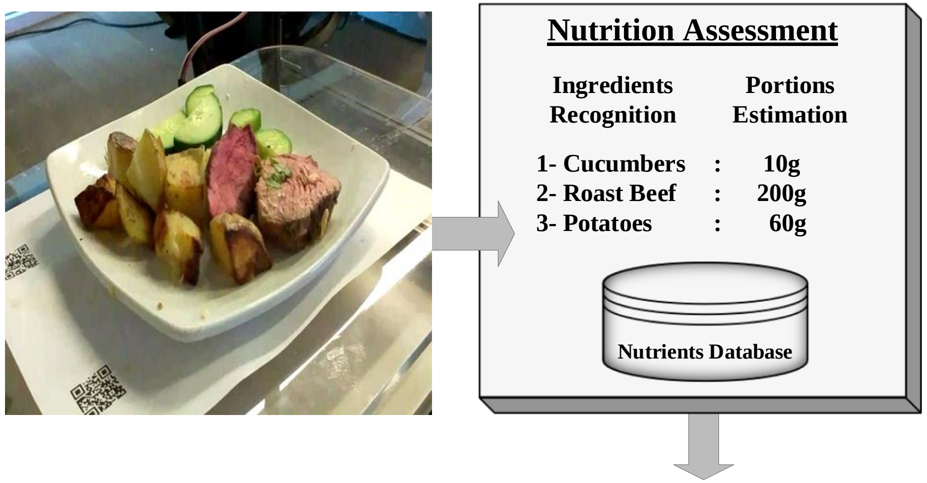
Introduction

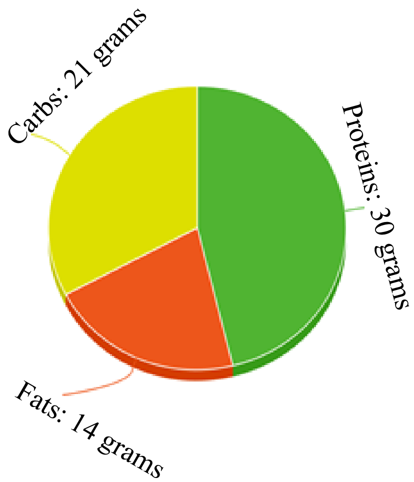
Automated Food Ingredient Recognition

1) Visual recognition of the food-items
2) Estimation of portion sizes
3) Database lookup service

Multi-Label Deep Learning

A deep neural network is trained to predict the presence of a multitude of ingredients or food-items. Generally, the final read-out layer consists of a fixed number of neurons, each representing a unique class label.





Nutrition Facts

Serving Size		270 g
Amount Per Serving		
Calories		**329**
		% Daily Values
Total Fat	13.69g	**18%**
Saturated Fat	4.036g	**20%**
Trans Fat	-	
Polyunsatrated Fat	1.118g	
Monounsatrated Fat	6.985g	
Cholestrol	73mg	**24%**
Sodium	483mg	**21%**
Total Carohydrate	21.22g	**8%**
Dietary Fiber	2.g	**9%**
Sugars	3.19g	
Protien	29.84g	

Background

Why Automated Diet Assessment?

Empowers communities to take control of their own health by providing them with insights and a precision guidance.

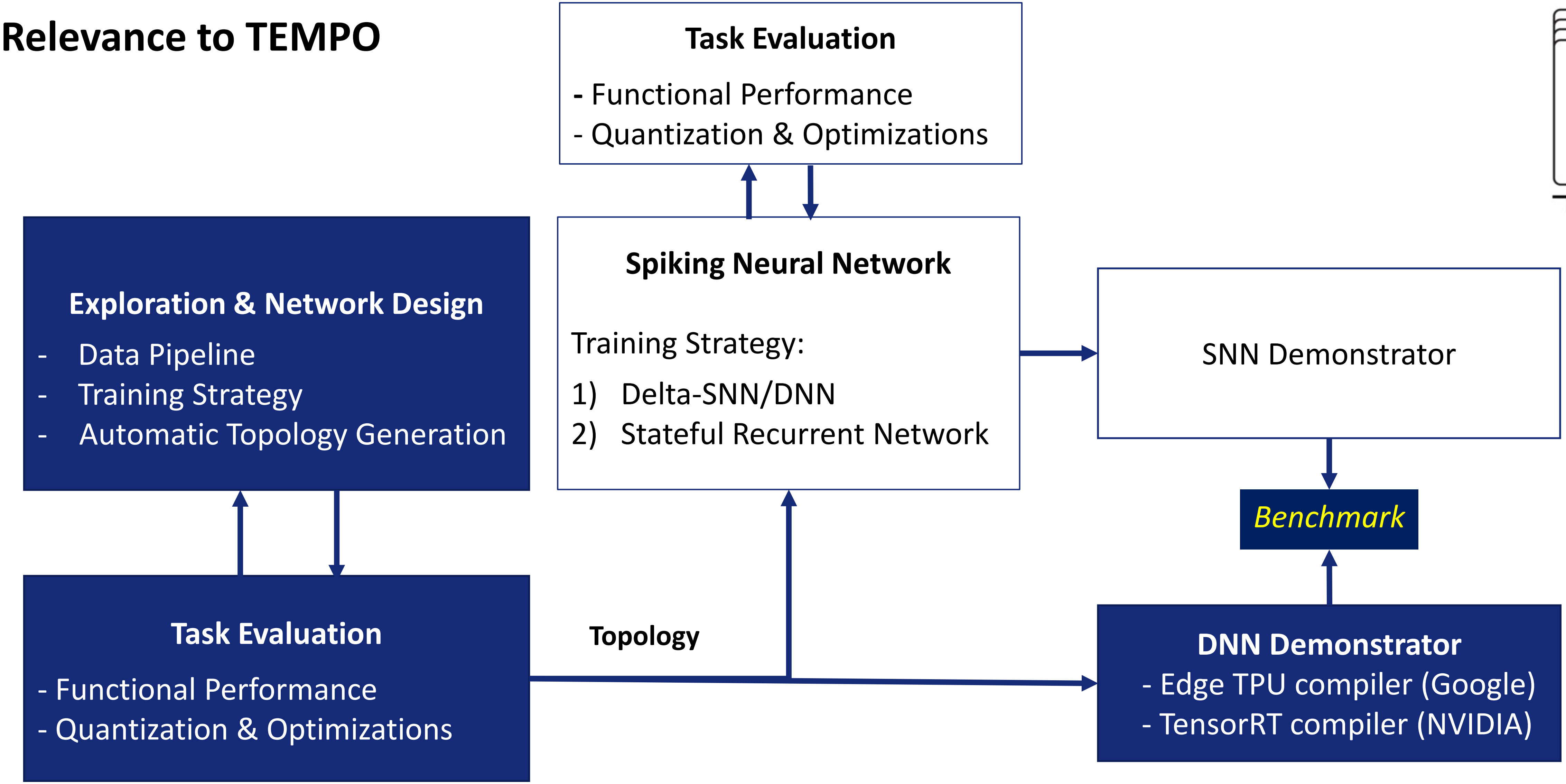
Enables better management of various nutrient-related health conditions and diseases at scale.

Besides providing an introspection ability to consumers, tracking dietary data is fundamental to understanding the diet and disease relationship.

Source: Lori Nedescu, MS RD CSSD, 'The Pros and Cons of Tracking Your Food', 2019.

Background

Relevance to TEMPO

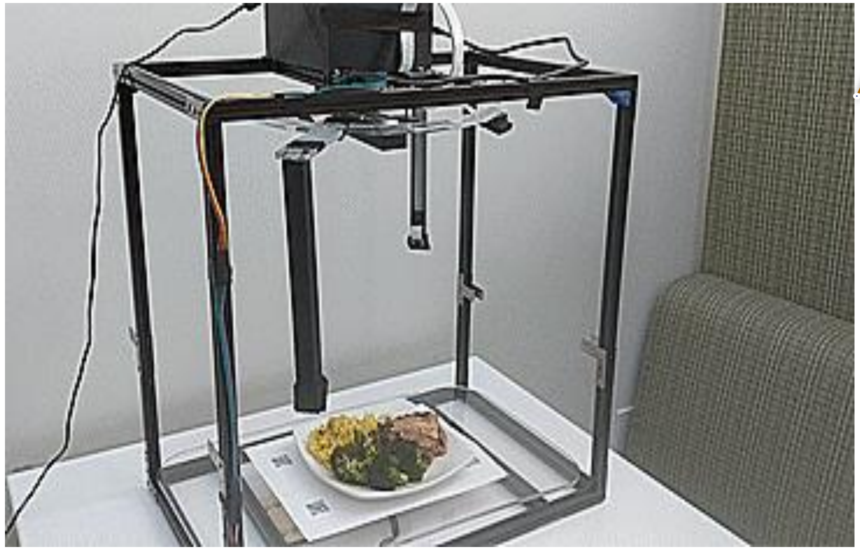


Research Goal

Focus of this Work

Evaluate state-of-the art deep learning modules for the task of ingredient recognition.

For training and evaluation, we employed `**Nutrition5k**` dataset. A dataset comprised of 5k unique dish images containing over 250 different ingredients.



Source: Thames, Quin, et al. "https://github.com/google-research-datasets/Nutrition5k. 2021.

Examples of dish images from the test dataset.

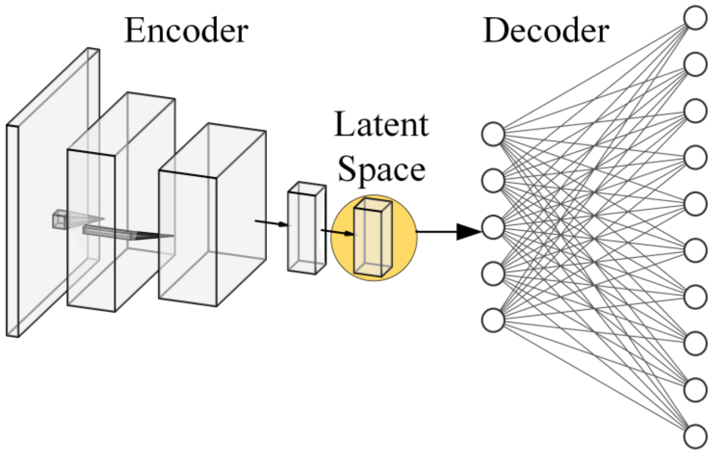
Design and Implementation

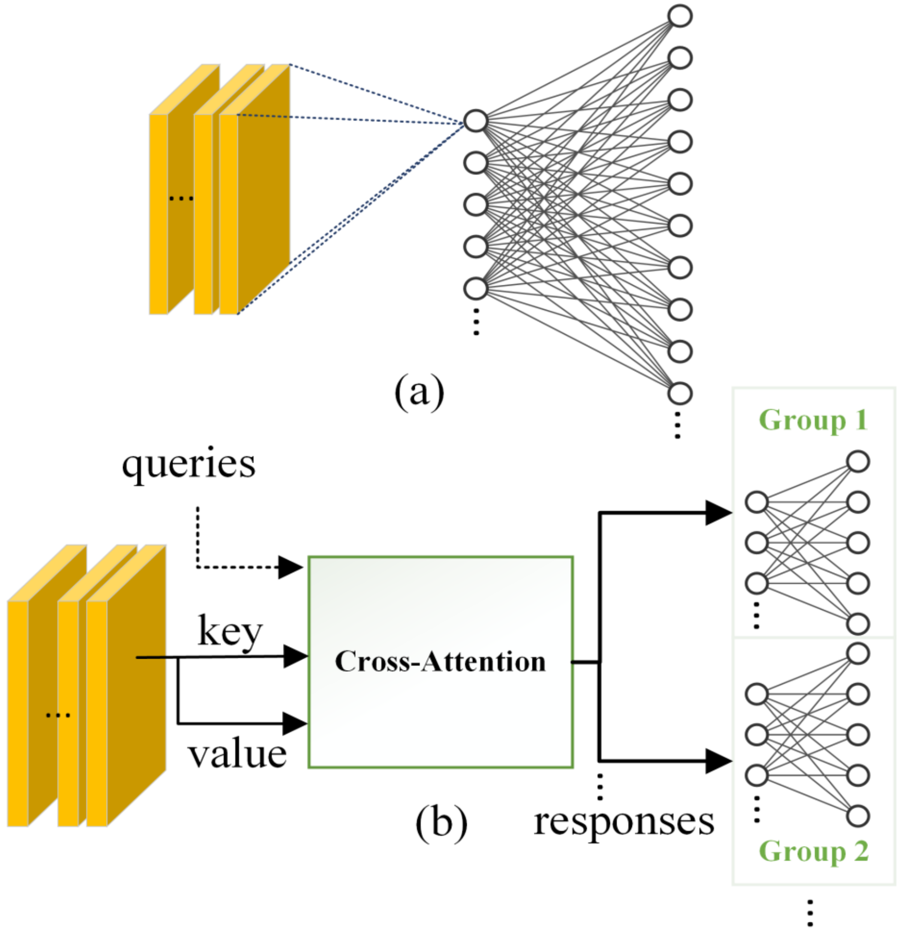
Design of Experiments and the Benchmark

All models consist of an encoder and a decoder component.

The various implementations evaluated for the encoder include, DenseNet, MobileNet, EfficientNet, Inception and Exception Network families.

Two distinct decoding schemes: global average pooling based and the other based on attention mechanism are evaluated and benchmarked.





Experimental Results

Results

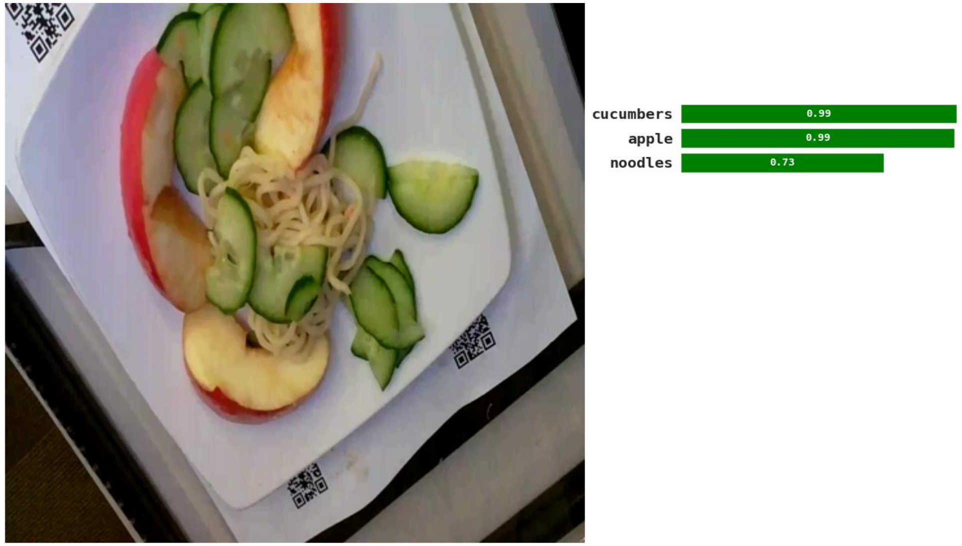
In the first set of experiments, we explored a global average pooling (GAP) based decoding, with different classification neural networks acting as the encoder.

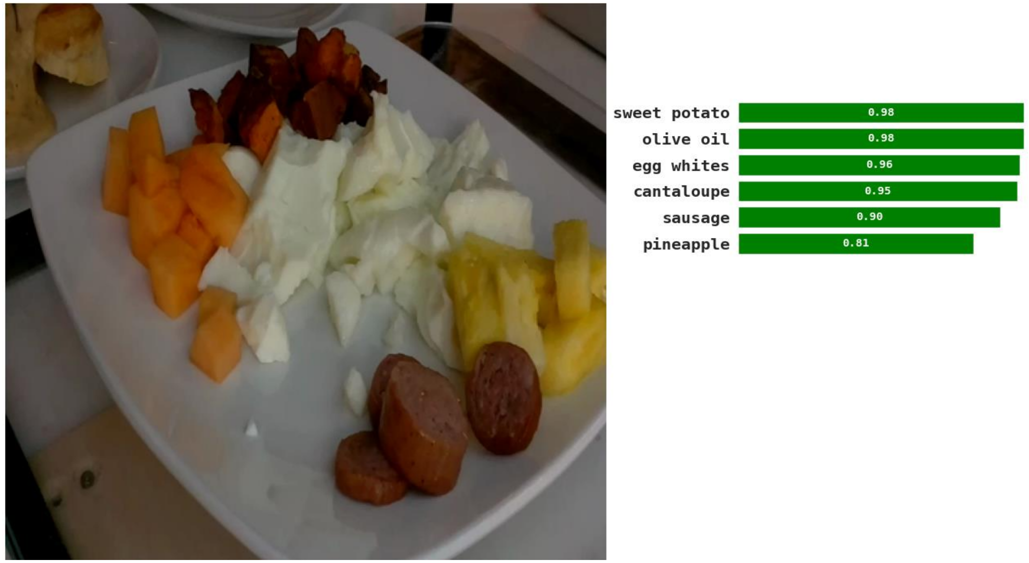
In the second set, we selected four encoders to couple with the ML-Decoder, based on their performance in the previous experiments and the compute specifications.

The GAP decoder performs best with Xception encoder reaching an accuracy of **78.4%** on the test set. The large performance differences, when using ML-decoder highlights its weakness as a drop-in replacement for GAP-based decoder.

Models		Performance	Compute	
Encoder	Decoder	mAP %	Operations (GFLOPs)	Parameters (MParams)
DenseNet121	GAP	75.6	22.6	7.6
DenseNet169		76.5	26.9	13.6
DenseNet201		74.7	34.3	19.4
MobileNetV1		72.4	4.5	3.8
MobileNetV2		74.5	**2.4**	**2.9**
EfficientNetB0		73.3	3.1	4.8
EfficientNetB1		71.9	4.6	7.3
EfficientNetB2		72.2	5.3	8.6
EfficientNetB3		71.5	7.8	11.6
EfficientNetB4		72.7	12.9	18.7
Xception		**78.4**	36.6	22.0
InceptionV3		72.8	26.4	23.0
MobileNetV2	ML-Decoder	68.0	3.4	9.3
EfficientNetB0		73.4	4.2	11.1
DenseNet169		67.9	28.0	19.9
Xception		70.0	38.0	28.5

Experimental Results





Discussion and Conclusion

A framework for automated diet assessment and encouraging results for image-based ingredient recognition using deep learning.

The current decoding schemes are unable to reliably extract the inter-label relationships. In future, we will further explore attention networks and graph neural networks that can exploit this information.

Refine the product and value proposition and pursue a Health-Suite Digital Platform offering.

Event Organisers

The Key Digital Technologies Joint Undertaking - the Public-Private Partnership for research, development and innovation – funds projects for assuring world-class expertise in these key enabling technologies, essential for Europe's competitive leadership in the era of the digital economy. KDT JU is the successor to the ECSEL JU programme. www.kdt-ju.europa.eu

ECSEL Joint Undertaking

The AI4DI project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826060. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the national authorities. www.ai4di.eu

AI4DI

The TEMPO project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826655. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, The Netherlands, Switzerland. www.tempo-ecsel.eu

TEMPO

The ANDANTE project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876925. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, The Netherlands, Portugal, Spain, Switzerland. www.andante-ai.eu

ANDANTE

Thank You

For your attention

@ Rameez.Ismail@philips.com