Multi-view Learning in Remote Sensing for Agricultural Applications

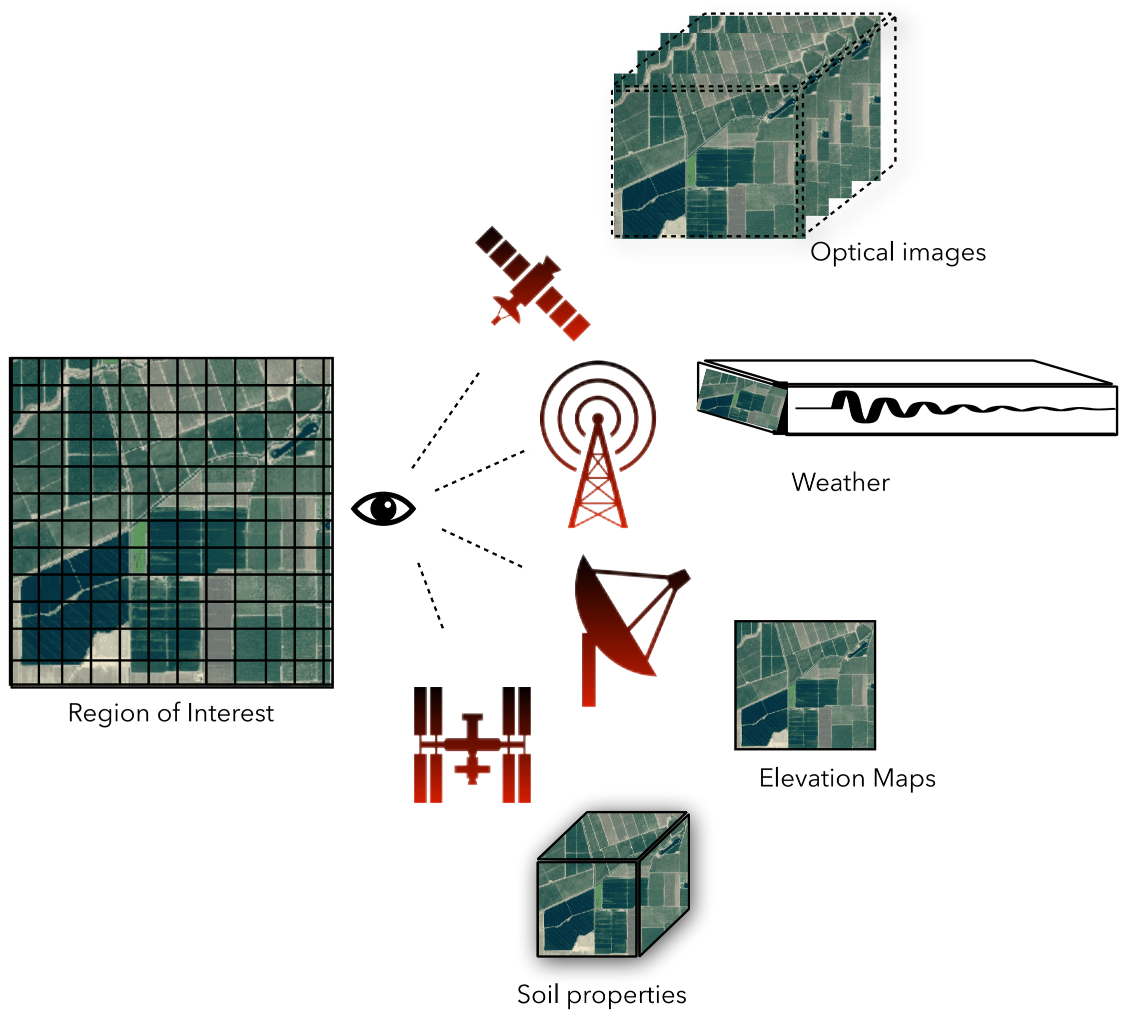
Francisco Mena[1,2], Mentor: Diego Arenas[2], Supervisor: Andreas Dengel[1,2]

[1]University of Kaiserslautern-Landau (RPTU) & [2]DFKI, Kaiserslautern, Germany

1. Introduction: Multi-view Learning (MVL), Remote Sensing (RS)

Precision agriculture aims to help farmers, industry stakeholders, and policymakers in optimizing agricultural practices and resource management. To aid this informed decision-making, machine learning provides predictive mapping models. Furthermore, the diverse and globally available RS sources could enhance the modeling of these tasks. However, due to the heterogeneous nature of RS data, such as varying spatial and temporal resolutions, combining this information is not a straightforward task. In RS literature, different fusion approaches have been tried in this MVL scenario. Here, two studies that assess fusion strategies in the agricultural field are presented.

2.1 Multi-view Remote Sensing Data



2.2 Case Studies: crops during growing season

CropHarvest [4] case study: **Binary crop classification** in a 10x10 m^2 pixel yearly (monthly sampled).

- **Input Views**: dynamic: multi-spectral optical image (12, 12), radar image (12, 2), weather (12, 2), and static: elevation map/DEM (2).

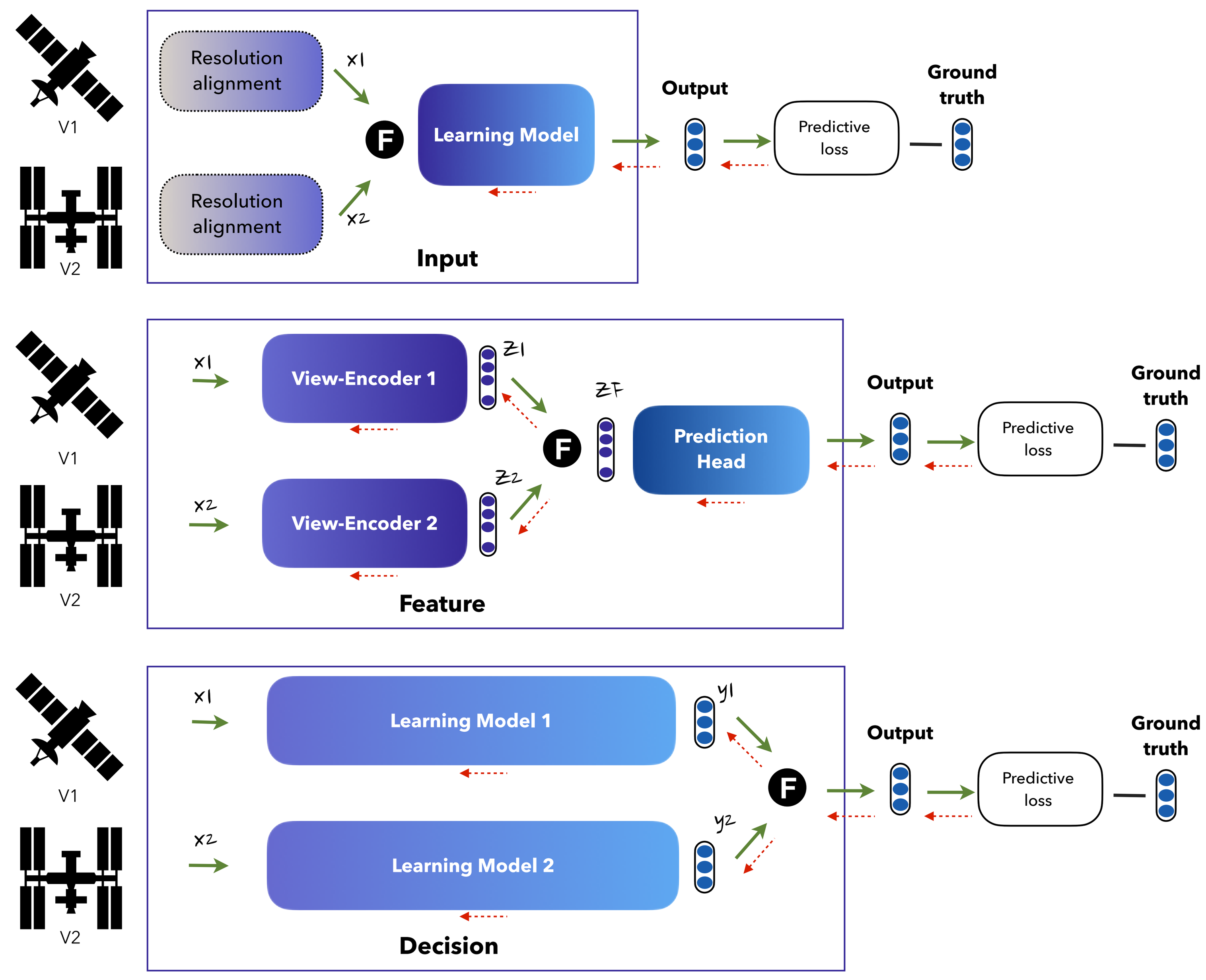
Data	Region	Training pixels (*positive*)	Testing pixels (*positive*)
KEN	Kenya	1319 (20.0%)	898 (64.0%)
TOG	Togo	1290 (55.0%)	306 (34.6%)

Private case study: **Crop yield prediction** in a 10x10 m^2 pixel time series.

- **Input views**: dynamic: multi-spectral optical image (∼5-day sampled, 12), weather (daily sampled, 4), and static: DEM (4), Soil properties (8).

Data	Region	Crop-type	# Fields	# Pixels
ARG-S	Argentina	soybean	190	∼ 1.4 M
GER-R	Germany	rapeseed	111	∼ 0.3 M

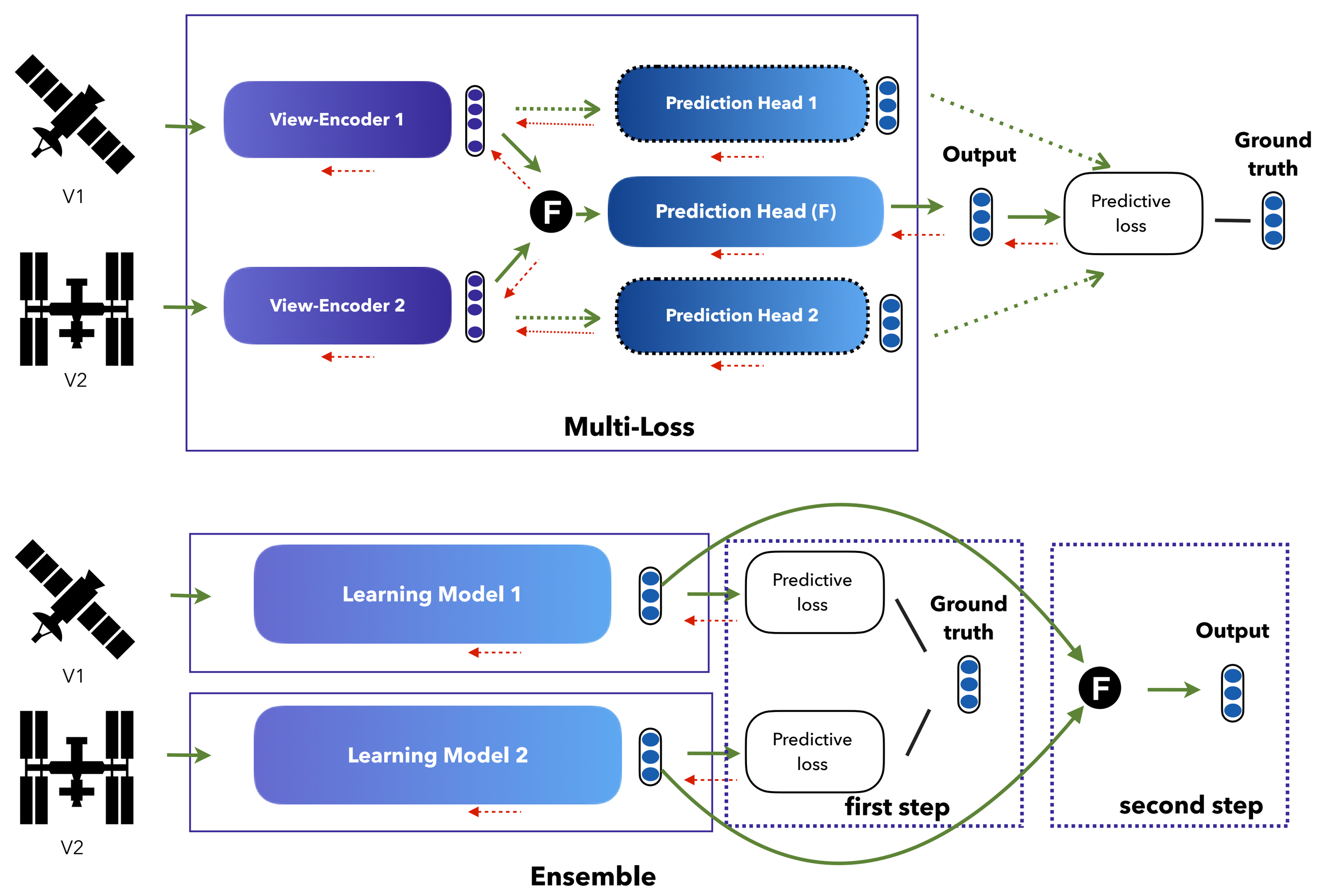
3. Fusion Strategies Compared (*inspired by [2, 3]*)



(A) Main strategies: Input, Feature, and Decision fusion.

$$z_F = \sum_v \alpha(z) \cdot z_v = \alpha_1(z_1, z_2) \cdot z_1 + \alpha_2(z_1, z_2) \cdot z_2$$

Gated: based on Feature Fusion and Gated Unit [1]



(B) Additional strategies: Gated, Multi-Loss, and Ensemble fusion.

4.1 Classification Case Study: Average Accuracy (AA)

Reference: Pixel Performance of (single) Weather or DEM ∼ 50% AA.

- One method does not fit all cases.
- Best performance fusion strategy depends on the testing region.
- 5% relative improvement compared to the best single view.
- Fusion closer to the output layer has higher prediction uncertainty.

Data	Method	AA	Entropy
KEN	*Radar*	63.0 ± 1.0	66.8 ± 2.9
	Input	61.3 ± 7.5	72.9 ± 6.1
	Feature	63.0 ± 5.1	73.9 ± 4.4
	Gated	$\mathbf{66.5 \pm 3.6}$	$\mathbf{71.6 \pm 5.8}$
	Multi-Loss	64.8 ± 5.2	73.4 ± 4.3
	Decision	57.5 ± 7.6	76.8 ± 7.6
	Ensemble	56.0 ± 5.3	83.5 ± 2.2
TOG	*Optical*	80.0 ± 1.3	88.6 ± 0.7
	Input	79.7 ± 1.5	54.3 ± 4.2
	Feature	79.9 ± 1.0	57.6 ± 4.1
	Gated	78.1 ± 1.8	$\mathbf{52.1 \pm 7.8}$
	Multi-Loss	78.2 ± 6.1	60.4 ± 12.2
	Decision	81.5 ± 1.6	63.7 ± 2.5
	Ensemble	$\mathbf{84.0 \pm 1.0}$	93.5 ± 0.6

`Published in IGARSS.` `arxiv: 2308.05407`

4.2 Regression Case Study: Coef. of determination (R^2)

Reference: Field Performance of Weather R^2 ∼ 0.20, DEM R^2 ∼ 0.0.

- Input* uses a subset of views as input.
- Best performance by Gated with all the views as input.
- 6 to 15% of relative improvement compared to (single) optical view.
- Pixel predictions much harder than field level.

`Partially published in IGARSS.` `arxiv: 2308.08948`

Data	Method	R^2 pixel	R^2 field
ARG-S	*Optical*	0.64	0.79
	Input*	0.65	0.82
	Feature	0.67	0.83
	Gated	**0.68**	**0.84**
	Multi-Loss	**0.68**	0.83
	Decision	0.67	0.83
GER-R	*Optical*	0.40	0.71
	Input*	0.45	0.78
	Feature	0.44	0.78
	Gated	**0.46**	**0.80**
	Multi-Loss	0.38	0.77
	Decision	0.48	0.76

5. Future Work

Optical images are affected by limiting or occluding factors like clouds and snow, or even sensor failure where the entire view is unavailable. Additional views could be helpful to supplement this missingness. My future steps consider designing a MVL model robust to missing views in RS.

References

[1] J. Arevalo et al. Gated multimodal networks. 2020.

[2] VSF Garnot et al. Multi-modal temporal attention models for crop mapping from satellite time series. 2022.

[3] F. Mena et al. Common practices and taxonomy in deep multi-view fusion for remote sensing applications. 2023.

[4] G. Tseng et al. CropHarvest: A global dataset for crop-type classification. 2021.