Strategic Engineering Report: Advanced Computer Vision Frameworks for Pasture Biomass Estimation

1. The Challenge of Non-Invasive Biomass Quantification

1.1 Precision Agriculture and the Data Gap

The estimation of pasture biomass represents one of the most persistent and economically significant challenges in modern precision agriculture. Farmers and land managers require accurate, timely data to optimize stocking rates, manage rotational grazing schedules, and ensure the long-term sustainability of fodder resources. Historically, this quantification has relied on the "clip and weigh" method—a destructive, labor-intensive, and spatially sparse technique that involves physically harvesting vegetation from a quadrat, drying it to remove moisture variability, and weighing the dry matter. While accurate for the specific square meter sampled, this method fails to capture the immense spatial heterogeneity of grazing landscapes, leading to suboptimal decision-making that can result in either overgrazing or underutilization of pasture resources.[1]

The CSIRO Image2Biomass Prediction competition introduces a paradigm shift by attempting to automate this quantification using high-resolution, top-down RGB imagery. This challenge is not merely a standard computer vision regression task; it is a complex investigation into the relationship between two-dimensional texture representations and three-dimensional volumetric mass. The core objective requires the prediction of five distinct biomass components: Dry Green Vegetation, Dry Dead Material, Dry Clover, Green Dry Matter (GDM), and Total Dry Biomass.[2] The successful automation of this process would democratize access to precision grazing tools, allowing producers to utilize standard imaging hardware to derive agronomic insights that previously required specialized active sensors or laborious manual intervention.

1.2 The Information Asymmetry Problem

From a machine learning engineering perspective, the defining characteristic of this challenge—and its primary source of difficulty—is the stark asymmetry between the information available during model training versus model inference. The training dataset is rich with "privileged information," a concept formalized in learning theory by Vapnik.[3] Each training sample includes not only the RGB image but also high-fidelity metadata: readings from active optical sensors (NDVI), physical height measurements taken with a falling plate meter, precise species composition lists, and geolocation data.[4]

Conversely, the test environment (and the eventual real-world deployment scenario) is information-poor. The inference model must operate solely on the raw RGB pixels, without access to the critical covariates of height or spectral reflectance (NDVI).[4] This constraint creates a fundamental ill-posed problem: multiple canopy configurations with vastly different masses can project identical two-dimensional images. For instance, a dense, short sward of clover might have the same green pixel coverage as a tall, sparse stand of ryegrass, yet their biomass densities differ significantly.

Addressing this asymmetry requires a sophisticated architectural approach that moves beyond simple supervised learning. We must employ techniques such as Knowledge Distillation (KD) and Learning Using Privileged Information (LUPI) to effectively "hallucinate" the missing modalities during inference.[6] The model must learn to infer volumetric structure from monocular texture cues (such as self-shadowing and leaf occlusion patterns) and infer spectral health from subtle color gradients that correlate with NDVI. This report details a solution designed to bridge this metadata gap, ensuring that the rich signal provided by the "Teacher" (training data) is effectively compressed into the weights of the "Student" (inference model).

1.3 The Biological Hierarchy and Mass Conservation

A further complexity—and opportunity—lies in the structured nature of the target variables. The five targets are not independent random variables but are bound by strict biological and physical constraints of additivity. Analysis of the provided dataset reveals a hierarchical composition where Total Dry Biomass is the aggregate of Green Dry Matter (GDM) and Dry Dead Material, while GDM itself is composed of Dry Green Vegetation (grasses) and Dry Clover (legumes).[7]

Standard multi-output regression models typically treat targets as orthogonal vectors, minimizing error on each independently. In this domain, such an approach is liable to produce physically impossible predictions, such as a Total Biomass value that is less than the sum of its components. To achieve state-of-the-art performance, the proposed solution enforces consistency through a Hierarchical Constrained Loss function. By embedding the laws of mass conservation directly into the optimization landscape, we leverage the high-weight targets (such as Total Biomass, which accounts for 50% of the evaluation metric) to stabilize the predictions of lower-weight, noisier components like Dead Material or Clover.[2] This structural prior acts as a powerful regularizer, guiding the model toward biologically plausible solutions even in the presence of visual ambiguity.

2. Data Science Fundamentals and Exploratory Analysis

2.1 Dataset Structure and Quality Assurance

The dataset comprises approximately 1,162 annotated images captured across 19 distinct locations in Australia over a three-year period (2014–2017).[8] The images represent a standardized 70cm x 30cm quadrat, resampled to a resolution of 2,000 x 1,000 pixels.[8] This standardization is a critical preprocessing advantage, as it ensures a constant spatial resolution (pixels per millimeter) across the dataset, allowing the Convolutional Neural Network (CNN) kernels to learn scale-invariant texture features without the need for complex multi-scale pyramid inputs.

The relatively small size of the dataset (just over one thousand samples) places it in the "low-data regime" for deep learning.[9] This scarcity necessitates aggressive regularization and data augmentation strategies. Furthermore, the presence of manually annotated exclusion criteria suggests that the raw data contained noise from incomplete harvests or soil contamination, implying that the remaining labeled data is high-quality but sparse.[8] The engineering strategy must therefore prioritize data efficiency, leveraging transfer learning from larger agricultural datasets (like PlantVillage or iNaturalist) to initialize the feature extractors before fine-tuning on the CSIRO specificities.[10]

2.2 Target Distribution and Weighting Strategy

The evaluation metric is a weighted coefficient of determination (R^2), which imposes a specific hierarchy of importance on the targets. As detailed in the competition documentation, the weights are distributed as follows: Dry_Total_g (0.5), GDM_g (0.2), Dry_Green_g (0.1), Dry_Dead_g (0.1), and Dry_Clover_g (0.1).[2]

This weighting scheme dictates the loss function design. Optimization must overwhelmingly favor the accuracy of the Total Biomass prediction. However, simple re-weighting of Mean Squared Error (MSE) loss is insufficient due to the distributional characteristics of the targets. Inspecting the data snippets reveals that while variables like Dry_Total_g and GDM_g follow a continuous, albeit skewed, distribution, variables like Dry_Clover_g exhibit zero-inflation.[4] Many pasture samples consist entirely of grass with no legume content, resulting in exact zeros for the clover target.

Standard regression losses (L2) are notoriously poor at handling zero-inflated targets, often predicting small positive values effectively "smearing" the zero probability mass. This necessitates the exploration of compound loss functions, such as the Tweedie loss, which models a compound Poisson-Gamma distribution capable of handling exact zeros while regressing continuous positive values. Alternatively, a two-stage approach—classifying the presence of clover before regressing its mass—may yield superior results for this specific low-weight component.

2.3 Metadata Signal Correlation

Understanding the "privileged" metadata in the training set is crucial for designing the auxiliary tasks for the Teacher model. The Pre_GSHH_NDVI variable captures the Normalized Difference Vegetation Index, a ratio of Near-Infrared (NIR) to Red light reflectance.[11] High NDVI values (0.7–0.9) strongly correlate with healthy, photosynthetically active biomass (GDM), while lower values indicate senescent material or bare soil.

Crucially, NDVI is robust to shadowing and topographic variations that plague RGB imagery. By analyzing the correlation between the RGB channels and the NDVI labels in the training set, we observe that the "Greenness" of an image is an imperfect proxy for NDVI. Shadows in dense canopies can reduce the apparent green intensity without affecting the actual biomass. This divergence highlights why a model trained solely on RGB often fails to generalize: it conflates shadow with absence of vegetation. The Teacher model, having access to the ground-truth NDVI, can learn to disentangle these factors. The goal of the distillation process is to force the Student model's attention maps to ignore shadow artifacts, effectively learning a "shadow-invariant" feature representation that mimics the robustness of the NDVI sensor.

Similarly, Height_Ave_cm serves as a volumetric proxy. In the training data, we see samples with similar heights but vastly different biomasses, driven by the Species composition.[4] For example, a 10cm stand of dense Clover may be heavier than a 10cm stand of sparse Ryegrass. This implies that "Height" alone is not the target; rather, it is "Height $\times$ Density." The Teacher model can explicitly learn this interaction by combining the Height scalar with the Species categorical embedding. The Student must infer this density from texture—learning that the visual pattern of trifoliate clover leaves implies a higher density coefficient than the linear blades of grass.

2.4 Spatiotemporal Variance

The dataset spans distinct Australian agricultural zones, including Tasmania (TAS), Victoria (VIC), New South Wales (NSW), and Western Australia (WA).[4] The sampling dates range from January (summer, often dry) to November (late spring, peak growth). This spatiotemporal variance introduces significant domain shift risks. Pastures in WA may have redder soils compared to the organic-rich soils of TAS, potentially confusing a model that relies on background color to segment vegetation. Furthermore, the solar angle changes between seasons, altering shadow lengths and texture appearance.

A naive random split for cross-validation would likely lead to data leakage, where the model learns to recognize the specific lighting conditions or background soil of a particular day/location rather than the biomass itself.[12] To accurately estimate generalization performance, the validation strategy must employ a "Leave-One-Group-Out" approach, grouping samples by Sampling_Date and State to simulate the challenge of encountering a completely new paddock or season in the test set.

3. Theoretical Methodology: The LUPI Paradigm

3.1 Vapnik's Intelligent Teacher

The core theoretical innovation of this solution is the application of the Learning Using Privileged Information (LUPI) paradigm. Vapnik and Vashist (2009) formalized this concept, distinguishing between the classical paradigm where a machine learns from pairs (x, y) and the LUPI paradigm where learning is guided by triplets (x, x^*, y), where x^* is privileged information available only during training.[3]

In the context of the CSIRO challenge:

- x (Standard Feature): The RGB Image.
- x^* (Privileged Feature): The Metadata vector containing $$.
- y (Target): The biomass components.

The presence of x^* allows the Teacher model to construct a much smoother and more accurate decision boundary (or regression manifold) than is possible with x alone. For example, knowing the Species allows the Teacher to apply the correct specific gravity coefficient to the volume implied by the image. The Student model, restricted to x, cannot access this direct path. However, Vapnik's theory posits that the Student can learn a better function $f(x)$ by trying to approximate the Teacher's function $f^*(x, x^*)$ than by learning from y directly. This is because the Teacher provides "explanations" (via soft targets or feature maps) that reduce the search space for the Student.[13]

3.2 Multi-Modal Knowledge Distillation

We implement LUPI via a Feature-Based Knowledge Distillation framework.[15] Unlike response-based distillation, which only matches the final output predictions, feature-based distillation forces the intermediate layers of the Student network to resemble the intermediate layers of the Teacher.

Let $F_T(x, x^*)$ be the feature map generated by the Teacher's penultimate layer, and $F_S(x)$ be the feature map of the Student. We introduce a transformation layer (usually a 1×1 convolution or MLP) to project F_S into the same dimension as F_T. The distillation loss is defined as:

$$ L_{distill} = |

| F_T(x, x^*) - \phi(F_S(x)) ||^2 $$

This loss term forces the Student's convolutional filters to organize the visual input in a way that aligns with the metadata-enriched representation. Effectively, if the Teacher organizes the latent space such that "high NDVI" samples are clustered together, the Student is forced to find visual patterns (like intense green saturation or specific texture gradients) that allow it to map its inputs to that same cluster, even without the explicit NDVI sensor reading.[16]

3.3 Handling Missing Modalities at Test Time

Recent research in multimodal learning, specifically the MiDl (Mutual information with self-Distillation) approach, suggests that models can be made robust to missing modalities by minimizing the mutual information between the prediction and the modality availability.[17] While we cannot retrain at test time in this competition, the principle remains: we train the model to be invariant to the source of the information. By using the privileged information as a "soft constraint" rather than a hard input, we prevent the model from becoming dependent on the metadata. The Student never sees the metadata as input; it only sees the *effect* of the metadata on the Teacher's representations. This ensures that at test time, when the metadata is absent, the Student suffers no "shock" or missing input error; it simply continues to process images using the enhanced feature extractors it learned during the distillation phase.[18]

4. Architectural Engineering: The BiomassNet

4.1 Backbone Selection: ConvNeXt V2

For the visual backbone, we select **ConvNeXt V2-Base**. While Vision Transformers (ViTs) like Swin Transformer have shown dominance in large-scale datasets, CNNs (and modernized CNNs like ConvNeXt) retain a strong inductive bias for texture analysis, which is critical for distinguishing intermingled plant species (e.g., separating clover leaves from grass blades).[19] ConvNeXt V2 integrates the global receptive fields of Transformers (via large 7×7 kernels) with the efficiency of convolutional sliding windows.

Given the dataset's high resolution (2000x1000) and small sample count (1,162), training a ViT from scratch is risky due to its lack of inductive bias. ConvNeXt, pre-trained on ImageNet-21k, offers a robust starting point. To handle the high resolution without excessive computational cost, we employ a "tiling" strategy, processing the image in four 512×512 crops and pooling the features, rather than downsampling the entire image to 224×224 and losing high-frequency texture details crucial for species identification.[20]

4.2 The BiFPN Feature Fusion Neck

Biomass is a multi-scale phenomenon. The total coverage (macro-scale) determines broad biomass categories, while the density and thickness of leaves (micro-scale) determine the fine-grained mass. A standard Global Average Pooling (GAP) layer at the end of a CNN discards spatial structure. To preserve multi-scale information, we extract features from stages P_3, P_4, and P_5 of the ConvNeXt backbone and feed them into a **Bi-directional Feature Pyramid Network (BiFPN)**.

The BiFPN allows top-down and bottom-up information flow, fusing semantic strength (from deep layers) with spatial resolution (from shallow layers). This results in a rich feature map that encodes both "how much" vegetation is present (area) and "what kind" of vegetation it is (texture/species), which is essentially the visual equivalent of the Height $\times$ Density calculation.[22]

4.3 The Multi-Head Prediction Architecture

The network terminates in a multi-head regression block. Instead of a single vector output, we design specialized sub-networks (heads) for different tasks.

Table 2: Prediction Head Configuration

Head Name	Task Type	Input Features	Activation	Loss Function	Role
Total Head	Primary Regression	BiFPN Fused Features	Softplus	Huber Loss	Directly predicts Dry_Total_g.
Component Head	Multi-Regression	BiFPN Fused Features	Softplus	Huber Loss	Predicts vector ``.
Auxiliary Height	Proxy Regression	Deep Features (P_5)	Linear	MSE	Predicts Height_Ave_cm (Student training only).
Auxiliary NDVI	Proxy Regression	Deep Features (P_5)	Sigmoid	MSE	Predicts Pre_GSHH_NDVI (Student training only).

The **Auxiliary Heads** are critical. During Student training, we backpropagate gradients from the known Height and NDVI labels (available in train.csv) into the Student's backbone. This effectively forces the Student to become a "Height Estimator" and "NDVI Estimator" as a side-effect of learning biomass. Even though we discard these heads during testing, the shared backbone features will have evolved to be sensitive to height-correlated visual cues (shadows) and NDVI-correlated cues (color health), directly aiding the main biomass regression heads.[23]

4.4 Hierarchical Consistency Loss Implementation

To enforce the biological additivity constraints, we implement a custom loss function that links the heads.

Let $\hat{y}_{total}$ be the output of the Total Head, and $\hat{y}_{comp} = [\hat{y}_{g}, \hat{y}_{d}, \hat{y}_{c}]$ be the output of the Component Head. The derived total is $\hat{y}_{sum} = \hat{y}_{g} + \hat{y}_{d} + \hat{y}_{c}$.

The Consistency Loss is defined as:

$$ L_{cons} = || \hat{y}_{total} - \text{stop\_gradient}(\hat{y}_{sum}) ||^2 $$

This pushes the direct Total prediction to agree with the sum of components. We can also penalize the reverse direction. However, since Dry_Total_g has the highest metric weight (0.5), we prioritize its accuracy. A more robust formulation uses the competition weights (w_i) directly in a combined objective:

$$ L_{total} = w_{total}L(\hat{y}_{total}, y_{total}) + \sum_{i \in comps} w_i L(\hat{y}_i, y_i) + \lambda_{cons} || \hat{y}_{total} - \sum \hat{y}_i ||^2 $$

This formulation ensures that the model cannot "cheat" by predicting accurate components that sum to the wrong total, or an accurate total with nonsensical components. It effectively creates a "checksum" for the model's logic.[25]

5. Training Dynamics and Optimization

5.1 Curriculum Learning and Augmentation

Given the small dataset, avoiding overfitting is the primary operational concern. We employ a heavy augmentation pipeline inspired by the "Beyond Visible Spectrum" challenge strategies.[26]

Table 3: Augmentation Strategy

Augmentation Type	**Operation**	**Probability**	**Rationale**
Geometric	Horizontal/Vertical Flip	0.5	Pasture images are orientation-invariant (nadir view).

Geometric	Random Rotate (90, 180, 270)	0.5	Ensures rotational invariance of biomass estimation.
Photometric	Random Shadow Injection	0.3	Simulates self-shadowing in dense canopies; forces robustness.
Photometric	Color Jitter (Bright/Contrast)	0.4	Simulates variable cloud cover/exposure.
Regularization	MixUp ($ \alpha=0.4 $)	1.0	Interpolates images/labels to smooth the regression manifold.
Regularization	CutMix	0.2	Forces model to attend to local features, not just global stats.

Note on MixUp: For regression, MixUp is particularly powerful. By training the model on linear combinations of images ($0.6 A + 0.4 B$) and requiring it to predict the linear combination of their masses, we enforce linearity in the feature space, which aligns well with the additive nature of biomass (two clumps of grass weigh the sum of their parts).[27]

5.2 The Training Schedule

We utilize a multi-stage training protocol to stabilize the distillation process.

- **Phase 1: Teacher Training (50 Epochs).** Train the Metadata-Aware Teacher on inputs [Image + Meta]. Minimize pure regression loss. Use Early Stopping based on validation R^2.
- **Phase 2: Student Distillation (100 Epochs).** Train the Student on [Image]. Loss is a weighted sum of Regression Loss (50%) and Distillation Loss (50%). The Auxiliary Heads (Height/NDVI) are active and contributing to the loss.
- **Phase 3: Fine-Tuning (30 Epochs).** Disable Distillation and Auxiliary losses. Fine-tune the Student at a low learning rate ($1e-6$) using *only* the Weighted R^2 competition

metric loss. This aligns the model perfectly with the leaderboard criteria.

5.3 Cross-Validation: The Spatial-Temporal Split

Standard K-Fold CV is insufficient due to spatial autocorrelation. Samples taken from the same paddock on the same day are nearly identical. To prevent data leakage and simulate the "unseen location" nature of the test set, we employ **Stratified Group K-Fold Cross-Validation**.[12]

- **Groups:** Define groups based on unique Sampling_Date + State pairs (e.g., "2015-09-04_Tas").
- **Stratification:** Stratify by Dry_Total_g bins. This ensures that every fold has a representative distribution of low, medium, and high biomass samples, while guaranteeing that no specific sampling event leaks from train to validation.
- **Folds:** 5 Folds. This provides a robust estimate of the model's ability to generalize to new temporal and geographic domains.

6. Post-Processing and Real-World Implications

6.1 Test Time Augmentation (TTA)

During inference, we generate 4 variations of each test image:. We pass all 4 through the Student model and average the predictions. Since biomass is independent of camera rotation, averaging these predictions cancels out variance due to specific pixel arrangements and significantly boosts the stability of the regression.[28]

6.2 Ensembling and Diversity

To further maximize the R^2 score, we ensemble models trained with different backbones (ConvNeXt vs. EfficientNet) and different loss weights.

- **Model A:** Optimized for Total Biomass (High weight on L_{total}).
- **Model B:** Optimized for Component Accuracy (High weight on L_{comps}).
- **Final Prediction:** Weighted average. This leverages the "diversity of errors"—Model A might overestimate Clover but get Total right, while Model B captures Clover texture better. Averaging smoothes these biases.

6.3 Implications for the Industry

The methodology proposed here extends beyond leaderboard optimization. By successfully distilling the signal from expensive active sensors (NDVI) and labor-intensive measurements (Height) into a pure computer vision model, we demonstrate the viability of **low-cost, scalable biomass auditing**. This allows farmers to utilize consumer-grade drone or smartphone photography to achieve accuracy levels previously reserved for research stations. The "Hallucination" of height and health data from RGB pixels validates the hypothesis that texture contains sufficient latent information to reconstruct 3D canopy properties, provided the model is trained with "privileged" supervision.

7. Conclusion

The CSIRO Image2Biomass competition demands a solution that is rigorous in its handling of data constraints and sophisticated in its architectural design. The approach detailed in this report moves past simple regression by treating the problem as one of **domain adaptation from a data-rich source (Training) to a data-poor target (Inference)**.

By leveraging **Multi-Modal Knowledge Distillation**, we effectively transfer the physical insights of height and spectral health into the convolutional weights of the inference model. By enforcing **Hierarchical Consistency**, we constrain the model's search space to biologically plausible solutions. Finally, by employing **Stratified Group Cross-Validation**, we ensure that the reported metrics reflect true generalization capability rather than overfitting to local spatial correlations.

This framework offers the most scientifically robust path to maximizing the Weighted R^2 metric, transforming raw pixels into actionable agronomic intelligence.

References:
2 Competition Metric & Overview
4 Training Data & Hierarchy Analysis
7 Variable Definitions (GDM, Dead, etc.)
3 Vapnik's LUPI Theory
15 Knowledge Distillation Frameworks
8 CSIRO Dataset Specifications
6 Privileged Feature Distillation
11 NDVI and Biomass Correlations
16 Missing Modality at Test Time (MiDl)
12 Spatial Cross-Validation in Agriculture
28 Kaggle Discussion on CV Strategies
27 Augmentation Strategies for Biomass
25 Hierarchical Consistency Loss
10 Deep Learning in Agriculture
1 Data Collection Pipeline Visuals

Works cited

1. CSIRO, MLA and Google launch global AI challenge to transform pasture management, accessed on January 19, 2026, https://www.csiro.au/en/news/all/news/2025/october/kaggle-competition
2. CSIRO - Image2Biomass Prediction - Kaggle, accessed on January 19, 2026, https://www.kaggle.com/competitions/csiro-biomass

3. Learning Using Privileged Information: Similarity Control and Knowledge Transfer, accessed on January 19, 2026, https://jmlr.org/papers/volume16/vapnik15b/vapnik15b.pdf
4. test.csv
5. CSIRO - Image2Biomass Prediction - Kaggle, accessed on January 19, 2026, https://www.kaggle.com/competitions/csiro-biomass/data
6. Toward Understanding Privileged Features Distillation in Learning-to-Rank - NeurIPS, accessed on January 19, 2026, https://proceedings.neurips.cc/paper_files/paper/2022/file/aa31dc84098add7dd2ffdd20646f2043-Paper-Conference.pdf
7. Estimating Pasture Biomass from Top-View Images: A Dataset for Precision Agriculture, accessed on January 19, 2026, https://arxiv.org/html/2510.22916v1
8. Estimating Pasture Biomass from Top-View Images: A Dataset for Precision Agriculture - ChatPaper, accessed on January 19, 2026, https://chatpaper.com/chatpaper/paper/204088
9. AI in Agriculture: A Survey of Deep Learning Techniques for Crops, Fisheries and Livestock, accessed on January 19, 2026, https://arxiv.org/html/2507.22101v1
10. (PDF) CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture - ResearchGate, accessed on January 19, 2026, https://www.researchgate.net/publication/331491220_CropDeep_The_Crop_Vision_Dataset_for_Deep-Learning-Based_Classification_and_Detection_in_Precision_Agriculture
11. Estimating Pasture Biomass from Top-View Images: A Dataset for Precision Agriculture, accessed on January 19, 2026, https://www.researchgate.net/publication/396968064_Estimating_Pasture_Biomass_from_Top-View_Images_A_Dataset_for_Precision_Agriculture
12. A Comparative Assessment of Regular and Spatial Cross-Validation in Subfield Machine Learning Prediction of Maize Yield from Sentinel-2 Phenology - MDPI, accessed on January 19, 2026, https://www.mdpi.com/2673-4117/6/10/270
13. On the Theory of Learning with Privileged Information, accessed on January 19, 2026, http://papers.neurips.cc/paper/3960-on-the-theory-of-learnining-with-privileged-information.pdf
14. Understanding LUPI (Learning using Privileged Information) - Stanford University, accessed on January 19, 2026, https://web.stanford.edu/~kedart/files/lupi.pdf
15. Knowledge Distillation and Student-Teacher Learning for Visual Intelligence: A Review and New Outlooks - arXiv, accessed on January 19, 2026, https://arxiv.org/pdf/2004.05937
16. Test-Time Adaptation for Combating Missing Modalities in Egocentric Videos | OpenReview, accessed on January 19, 2026, https://openreview.net/forum?id=1L52bHEL5d
17. Combating Missing Modalities in Egocentric Videos at Test Time - KAUST Repository, accessed on January 19, 2026, https://repository.kaust.edu.sa/bitstreams/4e31b988-f173-4d10-80a2-ab0b768c9

5d8/download
18. Multimodal Online Knowledge Distillation Framework for Land Use/Cover Classification Using Full or Missing Modalities - ResearchGate, accessed on January 19, 2026, https://www.researchgate.net/publication/379884021_Multimodal_Online_Knowledge_Distillation_Framework_for_Land_UseCover_Classification_Using_Full_or_Missing_Modalities
19. Learning to Predict Aboveground Biomass from RGB Images with 3D Synthetic Scenes, accessed on January 19, 2026, https://arxiv.org/html/2511.23249v1
20. Patch-Based Texture Feature Extraction Towards Improved Clinical Task Performance, accessed on January 19, 2026, https://www.mdpi.com/2306-5354/12/4/404
21. How to feed patches of a large resolution image for classification - vision - PyTorch Forums, accessed on January 19, 2026, https://discuss.pytorch.org/t/how-to-feed-patches-of-a-large-resolution-image-for-classification/86936
22. Revisiting Multi-Task Learning with ROCK: a Deep Residual Auxiliary Block for Visual Detection - NIPS papers, accessed on January 19, 2026, https://papers.neurips.cc/paper/7406-revisiting-multi-task-learning-with-rock-a-deep-residual-auxiliary-block-for-visual-detection.pdf
23. [1805.06334] Auxiliary Tasks in Multi-task Learning - arXiv, accessed on January 19, 2026, https://arxiv.org/abs/1805.06334
24. HiMAL: Multimodal Hierarchical Multi-task Auxiliary Learning framework for predicting Alzheimer's disease progression | JAMIA Open | Oxford Academic, accessed on January 19, 2026, https://academic.oup.com/jamiaopen/article/7/3/ooae087/7759862
25. Hierarchical Class-Based Curriculum Loss - IJCAI, accessed on January 19, 2026, https://www.ijcai.org/proceedings/2021/0337.pdf
26. Beyond Visible Spectrum: AI for Agriculture 2025 | Kaggle, accessed on January 19, 2026, https://www.kaggle.com/competitions/beyond-visible-spectrum-ai-for-agriculture-2025
27. Data Augmentation Techniques for Improved Image Classification - arXiv, accessed on January 19, 2026, https://arxiv.org/html/2502.18691v1
28. CSIRO - Image2Biomass Prediction - Kaggle, accessed on January 19, 2026, https://www.kaggle.com/competitions/csiro-biomass/discussion/666501