

Research and Implementation of Multi-Dataset Training for Image Classification with Discrepant Taxonomies

Master Thesis Presentation

Björn Buschhäuser

Technical Faculty, Bielefeld University

September 6, 2025

Outline

- 1 Introduction & Idea
- 2 Method Overview
- 3 Relationship Selection Methods
- 4 Synthetic Ground Truth & Domain Adaptation
- 5 Universal Model Training
- 6 Results
- 7 Conclusion

The Challenge: Limited Scope of Traditional Models

- Traditional image classification models are trained on specific datasets
- Each model recognizes only a predefined set of categories
- Multiple models needed for different domains = inefficient storage and deployment

Current Approaches:

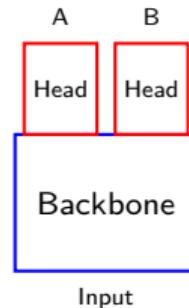
- **Transfer Learning:** Adapt pre-trained models to new tasks
- **Multi-head Architecture:** Shared backbone + task-specific heads
- **Multi-task Learning:** Train on multiple tasks simultaneously

Problem: Still requires separate models or heads for each domain

Our Idea: Universal Model vs. Multi-Head

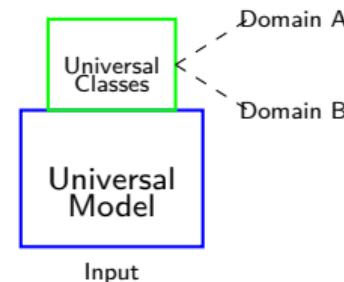
Multi-Head Approach

- Shared backbone
- Task-specific heads
- Automatic feature distillation
- Domain alignment challenges



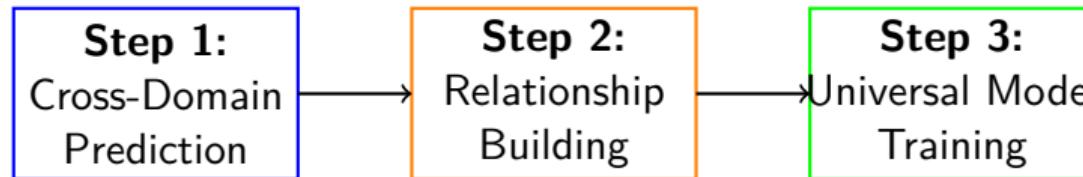
Universal Model Approach

- Single shared model
- Universal output layer
- Predefined concept mapping
- Static domain conversion



Key Insight: Build explicit concept-to-domain mappings to train a single universal model

Our Method in 3 Steps



- ① **Cross-Domain Prediction:** Train domain-specific models, run inference on other domains
- ② **Relationship Building:** Extract meaningful relationships from cross-domain predictions and create universal taxonomy
- ③ **Universal Model Training:** Create and train a single model using universal taxonomy

Cross-Domain Prediction Process

Goal: Discover relationships between classes from different datasets

Process:

- ① Train domain-specific models
- ② Run each model on images from *all other* domains
- ③ Create probability matrices $P_{ab}(i, j) = \text{probability of classifying class } c_i^a \text{ as class } c_j^b$

Example: CIFAR-100 model predicting on Caltech-101 images

- CIFAR-100 "automobile" → Caltech-101 "car_side" (high probability)
- CIFAR-100 "dog" → Caltech-101 "dalmatian" (moderate probability)
- CIFAR-100 "airplane" → Caltech-101 "butterfly" (low probability)

$$P_{ab}(i, j) = \frac{M_{ab}(i, j)}{\sum_{k=1}^{|C_a|} M_{ab}(i, k)} \quad (1)$$

Challenge: Selecting Relevant Relationships

Problems with raw probability matrices:

- Noisy predictions from imperfect models
- Unknown number of true relationships
- Different datasets have different scales of similarity

Solution: Develop & evaluate multiple relationship selection methods

Methods Evaluated:

- ① Naive Thresholding
- ② Most Common Foreign Prediction (MCFP)
- ③ Density Thresholding
- ④ Relationship Hypothesis

Evaluation Metrics:

- Edge Difference Ratio (EDR)
- Precision & Recall
- F1 Score

Relationship Selection Methods Explained

① Naive Thresholding:

$$\text{select_relationships}(P_{ab}) = \{(i,j) \mid P_{ab}(i,j) \geq t\} \quad (2)$$

② Most Common Foreign Prediction (MCFP):

$$\text{select_relationships}(P_{ab}) = \{(i,j) \mid j = \operatorname{argmax}_{j'} P_{ab}(i,j')\} \quad (3)$$

③ Density Thresholding: Select minimum relationships covering $p\%$ of probability mass

④ Relationship Hypothesis: Assumes relationships based on shared concepts should have equal probabilities. For each class, find optimal k relationships by minimizing:

$$\sum_{j=1}^k \left| X_i(j) - \frac{1}{k} \right| + \sum_{j=k+1}^{|C_b|} X_i(j) \quad (4)$$

where $X_i(j)$ are sorted probabilities in descending order.

Relationship Selection Results

Evaluation on synthetic datasets with known ground truth

Method	EDR	Precision	Recall	F1 Score
Naive Thresholding	0.463	0.675	0.889	0.767
MCFP	0.579	0.865	0.421	0.566
Density Thresholding	0.523	0.726	0.795	0.759
Relationship Hypothesis	0.493	0.746	0.823	0.783

Table: Global optimal performance across all synthetic dataset variants

Key Findings:

- **Naive thresholding** achieves best overall performance (lowest EDR)
- **MCFP** has highest precision but lowest recall
- Trade-off between capturing all relationships vs. avoiding false positives

The Need for Controlled Ground Truth

Problem: Real datasets lack clear inter-dataset relationships

- **WordNet:** Text-based relationships don't match visual similarities
- **Open Images:** Multi-label, automatically generated
- **Existing taxonomies:** Too domain-specific or hierarchical

Solution: Generate synthetic datasets with controlled relationships

- ① Define **atomic concepts** $\mathcal{U} = \{1, 2, \dots, n\}$
- ② Create synthetic classes as subsets: $c_j^i \subseteq \mathcal{U}$
- ③ Generate multiple domains by sampling concepts
- ④ Calculate relationships based on concept overlap

$$P_{i,j} = \frac{|c_i^A \cap c_j^B|}{|c_i^A|} + \frac{1 - \frac{|c_i^A \cap c_j^B|}{|c_i^A|}}{|C_B|} \quad (5)$$

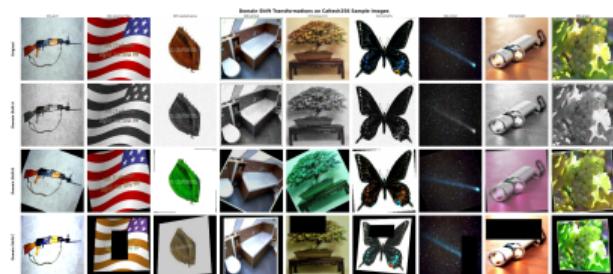
Domain-Shifted Synthetic Datasets

Problem: Original synthetic variants too easy (same underlying images)

Solution: Apply domain transformations to create realistic challenges

Transformations Applied:

- **Domain A:** Noisy grayscale
- **Domain B:** Rotation + blur
- **Domain C:** Random erasing + color jitter + perspective shifts



Results:

- Model accuracy drops by 10%
- More realistic cross-domain predictions
- Better evaluation of relationship selection methods

Example images from domain-shifted variants

Universal Model Architecture

Key Differences from Baseline Models:

Baseline Model:

- 6-layer FC classifier
- Dropout regularization
- One-hot targets
- Cross-entropy loss
- Domain-specific outputs

Universal Model:

- 2-layer FC classifier
- No dropout
- Probability distribution targets
- Cross-entropy for discrete distributions
- Universal class outputs

Target Generation: Convert domain labels to universal class distributions

$$\mathbf{t} = \hat{M}_i[j, :] \quad \text{where } \hat{M}_i(j, u) = \frac{M_i(j, u)}{\sum_{u'} M_i(j, u')} \quad (6)$$

Loss Function:

$$\mathcal{L} = - \sum_{u=1}^{|U|} \mathbf{t}(u) \log(\mathbf{p}(u)) \quad (7)$$

Multi-Domain Training Process

Training Procedure:

- ① Combine multiple datasets while preserving domain identity
- ② Each sample: $(\text{image}, (\text{domain_id}, \text{label})) \rightarrow (\text{image}, \text{universal_target})$
- ③ Train single model on unified dataset
- ④ Use domain-specific mapping matrices for target generation

Inference:

$$\mathbf{d}_i = M_i^T \mathbf{p} \quad (8)$$

$$\hat{c}_i = \text{argmax}(\mathbf{d}_i) \quad (9)$$

where \mathbf{p} are universal class predictions and \mathbf{d}_i are domain-specific predictions.

Key Challenge:

Validation loss increases while accuracy improves

- Solution: Monitor validation accuracy for checkpointing
- Caused by label smoothing effects and multi-target distributions

Universal Model Performance

Datasets: Caltech-101, Caltech-256, CIFAR-100

Model Type	Caltech-101	Caltech-256	CIFAR-100
Baseline (Single Domain)	0.828	0.798	0.734
Universal (MCFP)	0.845	0.812	0.728
Universal (Density)	0.834	0.809	0.741
Universal (Hypothesis)	0.831	0.801	0.735

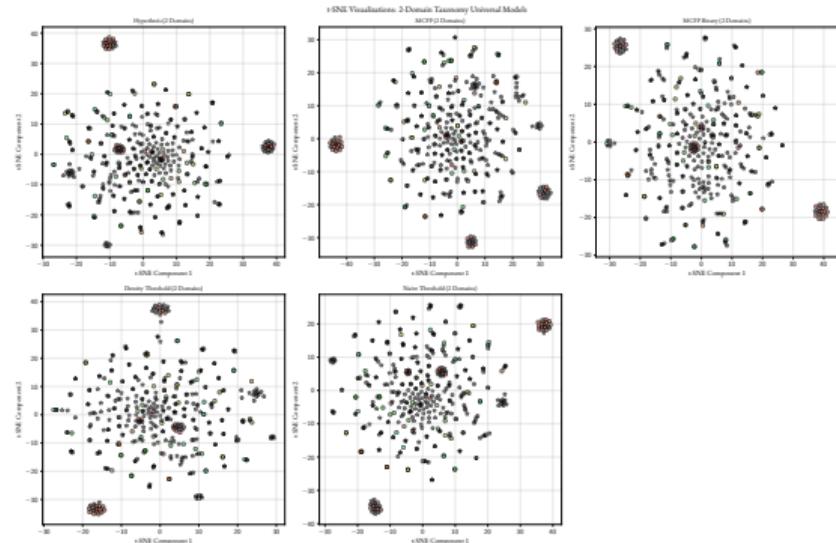
Table: Test accuracy comparison (2-domain: Caltech-101 + Caltech-256)

Key Findings:

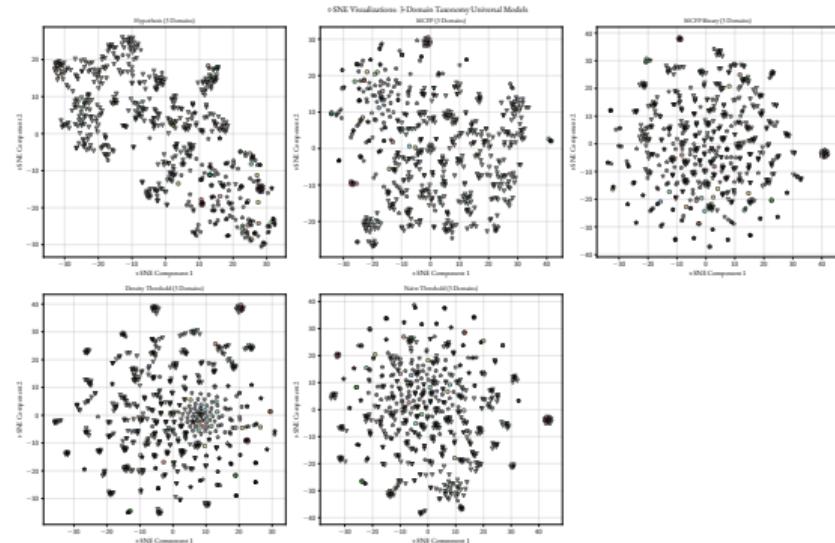
- Universal models **outperform** single-domain baselines
- Different relationship selection methods excel on different datasets
- No single method consistently optimal across all scenarios

Feature Visualization with t-SNE

2-Domain (Caltech-101 + 256)



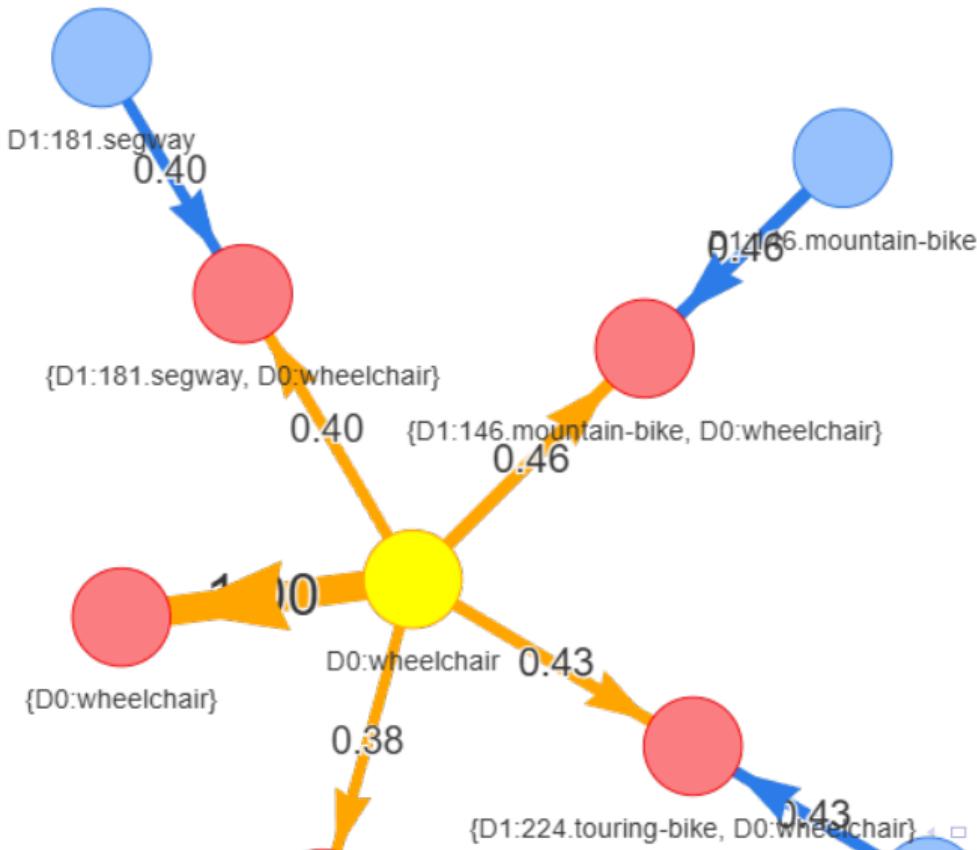
3-Domain (+ CIFAR-100)



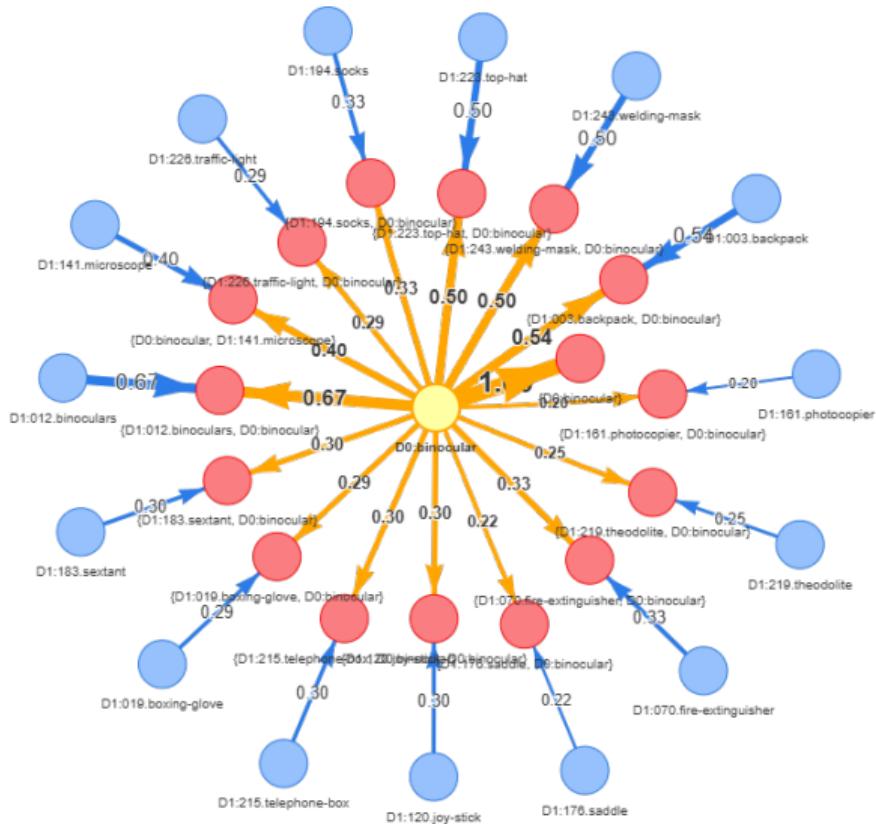
Observations:

- Mixed clusters indicate successful cross-domain feature learning
- Some domain-specific clusters for unique classes

Taxonomy Visualization Example



Challenges and Limitations



Key Contributions

- ① **Novel Universal Model Approach:** Single model for multiple domains without task-specific heads
- ② **Comprehensive Relationship Selection Methods:** Four different approaches with systematic evaluation
- ③ **Synthetic Ground Truth Framework:** Controlled evaluation environment with domain-shifted variants
- ④ **Cross-Domain Prediction Pipeline:** Complete methodology from raw datasets to universal models
- ⑤ **Performance Validation:** Universal models outperform single-domain baselines

Successful Aspects:

- Universal models achieve better accuracy than single-domain models
- Cross-domain prediction effectively captures semantic relationships
- Synthetic datasets provide valuable controlled evaluation environment
- Feature visualizations show meaningful cross-domain clustering

Challenges Identified:

- No single relationship selection method works optimally for all cases
- Gap between evaluation metrics and actual model performance
- Parameter selection requires careful tuning for each dataset combination
- Some false positive relationships in generated taxonomies

Future Work & Improvements

Immediate Improvements:

- Grid search for optimal relationship selection parameters
- Develop better correlation between evaluation metrics and model performance
- Explore different universal model architectures

Extensions:

- Scale to larger datasets (ImageNet, COCO)
- Apply to other vision tasks (object detection, segmentation)
- Investigate dynamic relationship adjustment during training
- Explore unsupervised relationship discovery methods

Real-World Applications:

- Multi-domain medical imaging
- Cross-dataset autonomous driving
- Unified content moderation systems

Problem: Traditional image classification limited to single domains

Solution: Universal model trained on cross-domain taxonomy

Method: Cross-domain prediction → Relationship selection → Universal training

Results: Universal models outperform single-domain baselines

Thank you for your attention!

Questions?