Improving Fine-Grained Vehicle Classification via Multitask Learning and Hierarchical Consistency

Gabriel E. Lima[*], Eduardo Santos[†,*], Eduil Nascimento Jr.[†], Rayson Laroca[‡,*], and David Menotti[*]
[*]{gelima,menotti}@inf.ufpr.br [†]{ed.santos,eduiljunior}@pm.pr.gov.br [‡]rayson@ppgia.pucpr.br

INTRODUCTION

Problem: Fine-grained vehicle classification often relies on compound labels, ignoring the hierarchical relationships between attributes. While multitask learning is a known alternative, its true benefits remain poorly understood, as prior studies are often restricted to limited experimental setups.

Objectives:

- To compare single-task vs. multitask learning across standard deep learning models.
- To assess a hierarchical regularization technique for enforcing semantic consistency.
- To establish a clear baseline to guide future research.

BACKGROUND

$$\mathcal{L}^{(k)}_{\text{MTL}} = \sum_{t\in\mathcal{T}} w_t^{(k)} \mathcal{L}_t^{(k)}$$

Weighted sum of individual task losses.

$$\mathcal{L}^{(k)}_{\text{GradNorm}} = \sum_{t\in\mathcal{T}} \left| G_t^{(k)} - \hat{G}_t^{(k)} \right|$$

Dynamically adjusts the weights using GradNorm.

$$\mathcal{L}^{(t_{d+1}|t_d)}_{\text{KL}} = D_{\text{KL}}\left(\mathbf{q}_x^{(t_{d+1}|t_d)} \parallel \mathbf{p}_x^{(t_{d+1})}\right)$$

KL Divergence penalty to prevent predictions that violate data's hierarchy.

DATASET

Source: Surveillance system.

Data: 24,945 images (16,308 unique vehicles).

Annotations: 26 makes, 136 models, and 14 vehicle types.

RESULTS

Classification Performance

Comparison of deep learning models for single-task, multitask, and multitask with hierarchical regularization setups. Hierarchical Consistency Error (HC-Err) is also reported, with (e3) achieving the largest reduction. Results are averaged over 10 runs, with standard deviation in parentheses. The best outcomes are shown in bold.

(e1) single-task learning — separate models are trained independently for each attribute.

Classification Model	Make			Model			Type			HC-Err ↓
	Mi-acc ↑	Ma-acc ↑	F1 ↑	Mi-acc ↑	Ma-acc ↑	F1 ↑	Mi-acc ↑	Ma-acc ↑	F1 ↑	
EfficientNet-V2 Small [31]	**94.43 (0.57)**	**84.99 (1.63)**	**86.36 (1.44)**	**90.91 (0.63)**	**86.16 (1.08)**	**87.25 (0.76)**	**96.12 (0.66)**	**88.97 (1.97)**	**90.23 (1.58)**	32.87 (1.74)
MobileNet-V3 Small [32]	91.27 (0.74)	77.99 (1.68)	80.19 (1.44)	86.52 (0.75)	79.16 (1.18)	80.90 (1.14)	95.15 (0.68)	85.61 (1.93)	87.41 (1.64)	36.84 (1.99)
ResNet-50 [33]	93.62 (0.51)	83.53 (1.73)	84.96 (1.08)	89.89 (0.87)	84.58 (1.21)	85.73 (0.79)	95.45 (0.61)	86.75 (2.34)	88.39 (1.90)	33.74 (1.80)
ResNet-101 [33]	93.80 (0.72)	83.49 (1.53)	85.01 (1.16)	90.20 (0.59)	84.87 (0.78)	86.09 (0.56)	94.69 (1.54)	84.64 (4.47)	85.92 (4.51)	**32.09 (1.25)**
ViT-B16 [34]	31.67 (1.68)	09.09 (3.43)	07.89 (3.04)	29.40 (1.71)	14.29 (3.58)	15.44 (3.51)	71.59 (1.52)	32.52 (5.17)	34.70 (5.09)	63.07 (2.33)
YOLOv11-nano-cls [35]	91.85 (0.99)	80.33 (2.81)	82.22 (2.17)	86.23 (1.38)	79.60 (1.94)	80.41 (1.73)	94.31 (0.80)	85.51 (2.54)	86.50 (2.16)	40.15 (3.99)
YOLOv11-small-cls [35]	93.07 (0.62)	82.68 (1.99)	84.36 (1.55)	87.53 (0.99)	81.33 (1.80)	82.47 (1.38)	95.20 (0.53)	87.00 (1.85)	88.18 (1.35)	38.38 (2.51)

(e2) Multitask learning — a single model jointly predicts all attributes.

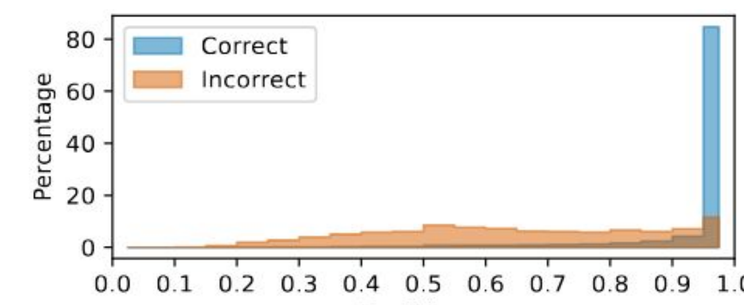
Classification Model	Make			Model			Type			HC-Err ↓
	Mi-acc ↑	Ma-acc ↑	F1 ↑	Mi-acc ↑	Ma-acc ↑	F1 ↑	Mi-acc ↑	Ma-acc ↑	F1 ↑	
EfficientNet-V2 Small [31]	**95.85 (0.54)**	**87.61 (1.16)**	**89.30 (1.20)**	**91.35 (0.82)**	**86.89 (1.50)**	**87.83 (1.32)**	**97.01 (0.55)**	**89.87 (2.06)**	**91.45 (1.80)**	**14.97 (1.73)**
MobileNet-V3 Small [32]	92.50 (0.64)	81.03 (1.83)	83.00 (1.53)	87.16 (0.83)	80.10 (1.29)	81.71 (1.04)	95.69 (0.68)	86.61 (1.90)	88.57 (1.69)	17.44 (1.68)
ResNet-50 [33]	95.14 (0.55)	86.49 (1.21)	87.82 (1.35)	90.40 (0.73)	84.72 (1.05)	86.26 (0.71)	96.62 (0.57)	88.61 (2.60)	90.46 (2.28)	16.75 (2.17)
ResNet-101 [33]	95.12 (0.51)	86.71 (1.01)	87.95 (1.10)	90.65 (0.66)	85.43 (0.95)	86.92 (0.79)	96.73 (0.63)	**90.16 (2.63)**	**91.61 (2.20)**	17.21 (1.39)
YOLOv11-nano-cls [35]	92.93 (0.59)	82.00 (1.73)	83.45 (1.44)	87.53 (0.84)	81.53 (1.15)	82.48 (0.75)	95.53 (0.76)	86.56 (2.51)	87.92 (2.22)	20.39 (1.75)
YOLOv11-small-cls [35]	93.72 (0.60)	84.08 (2.15)	85.52 (1.83)	88.47 (0.72)	83.07 (1.24)	83.71 (1.15)	95.86 (0.61)	86.55 (2.09)	88.40 (1.94)	19.18 (2.30)

(e3) Multitask with hierarchical regularization — KL-based penalties are applied to enforce consistency between related attribute predictions.

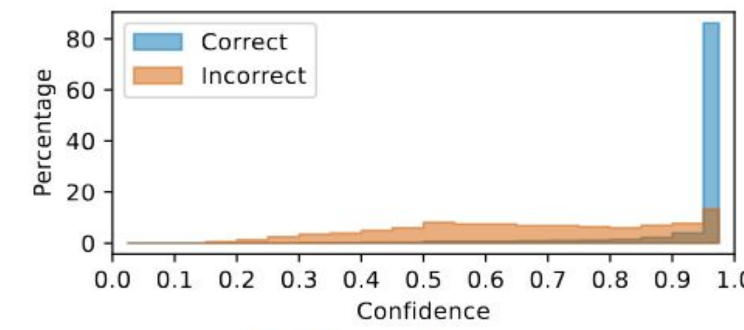
Classification Model	Make			Model			Type			HC-Err ↓
	Mi-acc ↑	Ma-acc ↑	F1 ↑	Mi-acc ↑	Ma-acc ↑	F1 ↑	Mi-acc ↑	Ma-acc ↑	F1 ↑	
EfficientNet-V2 Small [31]	**95.87 (0.55)**	**86.70 (2.18)**	**88.47 (1.70)**	**89.96 (1.36)**	**82.50 (1.36)**	**84.84 (1.29)**	**96.19 (0.69)**	84.62 (2.71)	86.95 (2.17)	**4.10 (0.89)**
MobileNet-V3 Small [32]	91.61 (0.84)	76.62 (2.87)	80.13 (2.33)	83.34 (0.71)	70.80 (1.32)	75.60 (1.15)	94.26 (0.72)	82.65 (1.68)	84.96 (1.18)	5.54 (0.68)
ResNet-50 [33]	95.16 (0.64)	84.92 (1.87)	87.30 (1.61)	88.82 (0.70)	79.67 (1.26)	83.60 (0.93)	**96.04 (0.67)**	85.21 (2.41)	**87.73 (2.12)**	**4.30 (0.86)**
ResNet-101 [33]	95.29 (0.70)	84.99 (1.77)	87.64 (1.48)	89.09 (0.80)	80.30 (1.27)	83.85 (0.94)	**96.22 (0.68)**	**85.90 (3.38)**	**88.36 (2.84)**	**4.41 (0.87)**
YOLOv11-nano-cls [35]	93.31 (0.72)	80.38 (2.22)	83.19 (1.92)	85.64 (0.86)	74.42 (1.40)	77.99 (1.21)	94.70 (0.75)	79.42 (2.10)	81.36 (2.00)	6.43 (0.88)
YOLOv11-small-cls [35]	94.52 (0.57)	83.76 (1.95)	86.16 (1.57)	87.95 (0.89)	78.92 (1.43)	81.84 (1.03)	95.35 (0.71)	79.93 (2.68)	81.72 (2.33)	5.71 (1.00)

Confidence Distribution

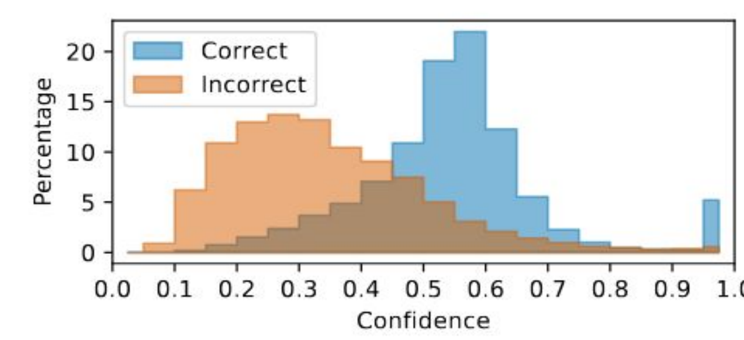
Vehicle model classification confidence across setups, using EfficientNet-V2 Small.



(a) Single-task setup.



(b) Multitask setup.



(c) Multitask + hierarchical regularization setup.

CONCLUSION

Key takeaways: Multitask learning generally improved classification performance. Hierarchical regularization offered a different benefit: it enhanced the model's consistency, even if it didn't always increase accuracy.

Next steps: Expand to finer-grained attributes (e.g. subtypes, sub-models), and find the optimal attribute combination for classification.

ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES), through the Programa de Excelência Acadêmica (PROEX) — Finance Code 001, and in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (# 315409/2023-1)