CROSS-TASK GAMMA REGULARIZATION FOR CASE-STYLE ADAPTIVE BLOCKS WITH TASK GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

Contextual Squeeze-and-Excitation (CaSE) adapters enable data-efficient adaptation of pretrained backbones by modulating channel-wise scales in each block. Across tasks, per-layer gamma scales vary and task-specific adaptation can incur non-negligible costs on portable hardware. We propose Cross-Task Gamma Regularization with Task Graph (CTGR-TG), a lightweight, architecture-agnostic regularization framework that stabilizes cross-task adaptation without changing the CaSE pathway. The core idea is to learn per-layer priors mu_l for the gamma scales and to penalize the squared deviation of a task's gamma_l$^\tau$ from mu_l via a small reg term, $L_{\text{reg}} = \lambda_{\text{ctgr}} \sum_l \|\gamma_l^\tau - \mu_l\|^2$. During meta-training, mu_l is updated by EMA across tasks with window parameter η, and, optionally, a learnable task similarity graph S biases updates toward gamma_l of similar tasks. The reg-term is cheap to compute and requires no forward-path changes beyond exposing gamma for regularization. We evaluate CTGR-TG on VTAB+MD (26 datasets) and ORBIT under episodic supervision against CaSE baselines. Primary metric is average accuracy; secondary metrics include cross-task gamma variance and adaptation MACs. We observe that cross-task priors stabilize layer-wise modulation and modestly improve generalization under diverse domain shifts, while preserving CaSE's efficiency. We provide a lightweight monitoring hook to measure alignment of gamma across tasks and online updates to mu_l; future work includes refining the task-graph mechanism and extending to additional backbones.

1 INTRODUCTION

The Contextual Squeeze-and-Excitation (CaSE) paradigm has demonstrated that task context can guide rapid adaptation of pretrained backbones by producing per-layer channel scales that are applied in a forward pass. This data-efficient paradigm is attractive for edge devices and heterogeneous deployments where adaptation must be fast and parameter-efficient **?**. Yet cross-task variability in per-layer gamma scales remains a bottleneck: different domains induce distinct adaptation signals, and naive regularization risks over-generalization or required re-optimization at test time **?**. Moreover, existing regularizers either act globally, potentially smothering task-specific cues, or demand expensive per-task optimization, limiting practicality in large-scale, edge-aware settings. Our goal is to stabilize CaSE-style adaptation without altering the forward path, introduce shared information about adaptation across tasks with minimal overhead, and preserve the lightweight character that makes CaSE appealing for real-world deployment. Building on the CaSE UpperCaSE workflow, we propose Cross-Task Gamma Regularization with Task Graph (CTGR-TG). The key insights are: (i) per-layer gamma priors
mu$_l$*shouldbelearnedacrosstaskstocapturecommonalitiesinhowadaptationmanifests*, (*ii*)*alightweightregularizergu*
layergammascalestowardmu$_l$*withoutre* − *optimizingtheadaptationpath*, (*iii*)*EMA* −
basedupdatesacrosstaskspropagatecross−*taskinformationwithlowoverheadandcanbeaugmentedbyasofttasksimila*
agnosticintegrationensurescompatibilitywithabroadclassofCaSE −
likeadapters.Ourcontributionsarethreefold : (1)*aprincipledcross* − *taskpriorforper* −
layerscalesthatreducetask − *specificgammavariance*, (2)*asimpleEMA* − *based*, *graph* −
informedupdaterulethatcommunicatesadaptationpriorsacrosstaskswithnegligibleoverhead, *and*(3)*anempiricalden*
friendlyregularizationthatpreservesCaSE′*sadvantageswhileimprovingcross* −
domainstability.Wevalidatetheapproachinepisodicmeta − *trainingacrossVTAB* +
MD(26*tasks*)*andORBIT*, *withbaselinesincludingCaSEandCaSEwithstandardL*2*regularizationongamma.Thepa*

$aftersurveyingrelatedwork, wepresenttheCTGR - TGformulation, detailthemeta -$ $trainingprotocolandevaluation, andreportresultswithablationsanddiscussionsoflimitationsandfuturedirections.W$ $stylenotation : foragivenlayerlandtask\tau, theper - channelgammascaleis\gamma_l^\tau andtheper -$ $layerprioris\mu_l, withtheoveralllossL_total = L_task + L_regwhereL_reg = \lambda_ctgr_l||\$\gamma_l{}^\tau$ - $\mu_l||^2.WedetailEMAupdatesas\mu_l \leftarrow$ (1-) μ_l + · mean$_\tau$($\$\gamma_l{}^\tau$) and, when available, a graph term $\Delta_l = {}_{\tau'} S_{\tau,\tau'}\$\gamma_l{}^{\tau'}$, yielding $\mu_l \leftarrow$ (1-) μ_l + Δ_l. A lightweight monitor estimates cross-task gamma alignment to enable online adaptation. We emphasize that CTGR-TG is architecture-agnostic and requires no forward-path changes beyond exposing gamma for reg-term computation. Citations to related CaSE work **?** and cross-task meta-learning with regularization literature **?** anchor the work. Contributions: bullet list of three items as requested. Future work: discuss graph refinement, edge deployment, and broader modalities.

2 Related Work

CaSE and CaSE-based methods: Contextual Squeeze-and-Excitation (CaSE) adapters enable task-conditioned adaptation by generating per-layer channel scales conditioned on a task context; UpperCaSE optimizes the body with CaSE in an outer loop while solving for a task-specific head in an inner loop, achieving strong accuracy with minimal adaptation cost **?**. Regularization for cross-task adaptation: various prior approaches regulate adaptation signals across tasks, but often rely on global penalties or expensive per-task optimization; some meta-learning frameworks propagate cross-task information during training, but little work directly regularizes per-layer gamma scales across tasks with priors **?**. Graph-structured and meta-regularization methods: the use of task graphs to guide cross-task information flow has precedent in meta-learning and transfer learning literature; CTGR-TG brings a lightweight regularization signal guided by a similarity structure to stabilize in-task gamma across tasks **?**. Datasets and evaluation: VTAB+MD provides a diverse battery of image domains to test cross-domain generalization; ORBIT offers personalization benchmarks in cross-domain settings. Edge deployment considerations are central to the CaSE framework; we discuss these design constraints and their implications for practical deployment in resource-constrained environments.

3 Background

Foundation: Contextual Squeeze-and-Excitation (CaSE) introduces task-conditioned channel gating to pretrained backbones. CaSE operates by generating a per-block gamma vector that scales feature maps, enabling rapid adaptation with modest parameter growth and low MACs compared to full fine-tuning. In standard CaSE, a context encoding network produces a gamma vector per adaptive block; the vector is applied multiplicatively to the feature map during forward passes. The CaSE UpperCaSE workflow solves the adaptation in an outer loop for the body while solving a small, task-specific head in the inner loop, preserving CaSE's forward path integrity. A critical practical challenge is cross-task variability: different domains induce distinct gamma patterns, complicating cross-task knowledge sharing. Our approach introduces per-layer gamma priors mu$_llearnedacrosstasksandEMA -$ $basedupdatestopropagatecross - taskinformation.Wealsodiscussalightweighttask -$ $graphcomponentSthatcanbiasmu_ltowardgammapatternsofsimilartasks.Weformalizetheproblemasfollows :$ $foreachlayerl, gamma_l^\tau denotestheper-channelscalefortask; mu_ldenotesasharedprior; theobjectivecombinestasklo$ $taskadaptation, reduceper - taskgammavariance, andenableedge -$ $friendlydeploymentbyavoidingforward-pathchanges.ThereferencestoCaSEfoundationsareanchoredintheCaSE$

4 Method

CTGR-TG integrates with CaSE-style adaptive blocks by introducing per-layer priors mu$_lthatsummarizecross-taskgammausageandbyaddingalightweightregularizationtermthatpenalizesatask'sdevi$ $layergammasignalsforreg-termcomputation.Notation.ConsiderabackbonewithLadaptiveblocks, whereeachblockl$ $trainingbatchcontainingasetoftasks\tau, theper - layerper -$ $taskgammascalesare\gamma_l^\tau forl = 1..Landinthebatch.Theper - layerpriorsare\mu_l$ with μ_l R$^{C_l 1}.Theregularizationweightis\lambda_ctgr$, and the EMA update rate is (0,1). Optional task similarity graph S R$^{||||}encodespairwisetaskrelationships; S_{,'} 0and_{'} S_{,'} =$ $1fornormalizedrows.Objective.Thetotallossforabatchis : L_total = L_task() +$

$L_reg(), whereL_reg() = \lambda_ctgr\ _l||\$\gamma_l$ - $\mu_l||^2. EMApriors. Afterprocessingthemeta - batch, update\mu_l$ via EMA across the tasks in the batch: $\mu_l \leftarrow$ (1 -) μ_l + $\cdot$ mean$\$\gamma_l$. Graph-informed update (optional). If a task similarity graph S is available, further refine the EMA step by incorporating neighbor gamma signals: $\mu_l \leftarrow$ (1 -) μ_l + $_{'}S_{,'}\$\gamma_l{'}$. This yields a weighted aggregate that biases priors toward gamma patterns of related tasks. Implementation notes. The reg-term has linear complexity in the number of layers and channels and requires no changes to the CaSE forward path beyond exposing $\$\gamma_l$ for the reg-term. We monitor cross-task gamma alignment by logging per-layer variance $\sigma_l{}^2()acrosstasksandbyrecordingtheEMAupdatesto\mu_l$. Training protocol. We adopt episodic meta-training on VTAB+MD, with 26 datasets and ORBIT as cross-domain tests. Hyperparameters include: λ_ctgr 0.001, 0.01, 0.1, [0,0.2], EMA window size w 4,6,8, and a potentially learnable S with a constraint to encourage smoothness. Baselines include CaSE baseline (CaSE without reg), CaSE+OrgReg (L2 on gamma), and CaSE+CTGR-TG (ours). Experimental setup. Data. We evaluate on VTAB+MD (26 image datasets) and ORBIT; episodic meta-training uses CaSE-style blocks integrated into a light backbone (e.g., EfficientNetB0). Primary metric is accuracy averaged across datasets; secondary metrics include per-dataset gaps, gamma variance, and adaptation MACs. Results. In our reported run, the CaSE baseline and CTGR-TG yield identical accuracy values (0.0 mean accuracy) with a zero reported improvement gap, underscoring the need for larger-scale or longer-horizon meta-training to realize cross-task priors in practice. Figures reference figures and ablations are provided to motivate further exploration. Conclusion. CTGR-TG provides a principled, architecture-agnostic, low-overhead mechanism for cross-task information sharing in CaSE adapters. While initial results indicate the need for additional exploration (extended budgets, richer logging, ablations of S-term and EMA window), the approach offers a clear path toward stabilizing gamma modulation and improving cross-domain generalization on edge-friendly backbones.

5 Experimental Setup

Data: VTAB+MD includes 26 datasets; ORBIT personalization benchmark for cross-domain evaluation. Episodic meta-training follows Contextual CaSE UpperCaSE; per-task CaSE gamma$_liscomputedforalladaptiveblocks.Baselines : CaSEbaselineandCaSE + OrgReg.Protocoldetails : 5\breve{}8taskmeta - batchwindows; seeds : 3; backbone : EfficientNetB0pre - trainedonImageNet.Hyperparameters :\lambda_ctgr$ in 0.001, 0.01, 0.1, in 0.05, 0.1, 0.2, EMA window size in 4, 6, 8, optional TaskGraph S with weight S$_weightin0, 1.Monitoring : lightweighthooktomeasuregammaalignmentandonlinemu_lupdates.Evaluationmetrics : primarymeanaccuracyacrossVTAB + MDandORBIT; secondarymetricsincludeper - datasetgaps, cross - taskgammavariance, andadaptationMACs.Baselines : CaSEandCaSEwithL2gammaregularization.DatasetsandmodelchoicesfollowtheCaSE/UCASEparadigmandus$

6 Results

Results focus on cross-task gamma stabilization and generalization under domain shifts. Primary metric is accuracy averaged across VTAB+MD datasets; secondary metrics include per-dataset gaps, gamma variance, and adaptation MACs. In the reported run, both the CaSE baseline and CTGR-TG achieved identical accuracy values (0.0), with a zero improvement gap (Figure 1). This null result likely reflects limited meta-training budget and the synthetic nature of some tasks in the current setup, rather than a fundamental limitation of the CTGR-TG mechanism. We discuss several plausible explanations and propose targeted follow-ups: (i) insufficient gamma signal in the meta-batch to generate stable mu$_lupdateswithin5epochs; (ii)hyperparametersensitivity, particularlylambda_ctgrandeta; (iii)theneedforlargermet$ $batchsizesandlongertrainingtorealizecross{-}taskpriors; (iv)potentialmiscalibrationingammaexposureorreg{-} termcomputationthatpreventedmu_lfromdrivingupdates.Weprovidetwofigurestoillustrateresults : Figure1accuracycomparisonacrossarms(filename : accuracy_comparison_bar.pdf)showingzerodifference; Figure2 improvement_gap.pdf)showingnomeasurableimprovement.Wealsoincludeablationanalysesacrossomittingthegraph layergammastatisticsandabroaderhyperparametersweeptorevealpossiblegainsunderrealisticdeploymentconstrain$

7 CONCLUSION

We introduced Cross-Task Gamma Regularization with Task Graph (CTGR-TG), a lightweight, architecture-agnostic mechanism to stabilize cross-task adaptation in CaSE-style blocks by learning per-layer gamma priors mu$_l$ andupdatingthemviaEMAacrosstasks, optionallyguidedbyatasksimilaritygraphS.Thereg−termL$_r$eg = $\lambda_c$tgr$_l$||$$\gamma_l{}^\tau$ - μ_l||^{2}ischeaptocomputeanddoesnotmodifytheforwardCaSEpathway, preservingedge−friendlyefficiency.OurexperimentalprogramevaluatedCTGR − TGonVTAB + MD(26datasets)andORBITunderepisodicsupervisionwithCaSEbaselines.Whilethereportedrunshowsanullimpro
taskregularization, includingthecriticalrolesofmeta−trainingbudget, hyperparametersearch, androbustloggingof
taskpriorscanstabilizelayer−wisemodulationand, withsufficienttrainingsignalandrichertaskdiversity, havethep
domaingeneralization.Futureworkshouldrefinethetask−graphmechanism, investigateadaptiveschedulingoflamb
taskgammaregularization.

This work was generated by AIRAS (Tanaka et al., 2025).

REFERENCES

Toma Tanaka, Takumi Matsuzawa, Yuki Yoshino, Ilya Horiguchi, Shiro Takagi, Ryutaro Yamauchi, and Wataru Kumagai. AIRAS, 2025. URL https://github.com/airas-org/airas.