Deep Hidden Analysis: A Statistical Framework to Prune Feature Maps

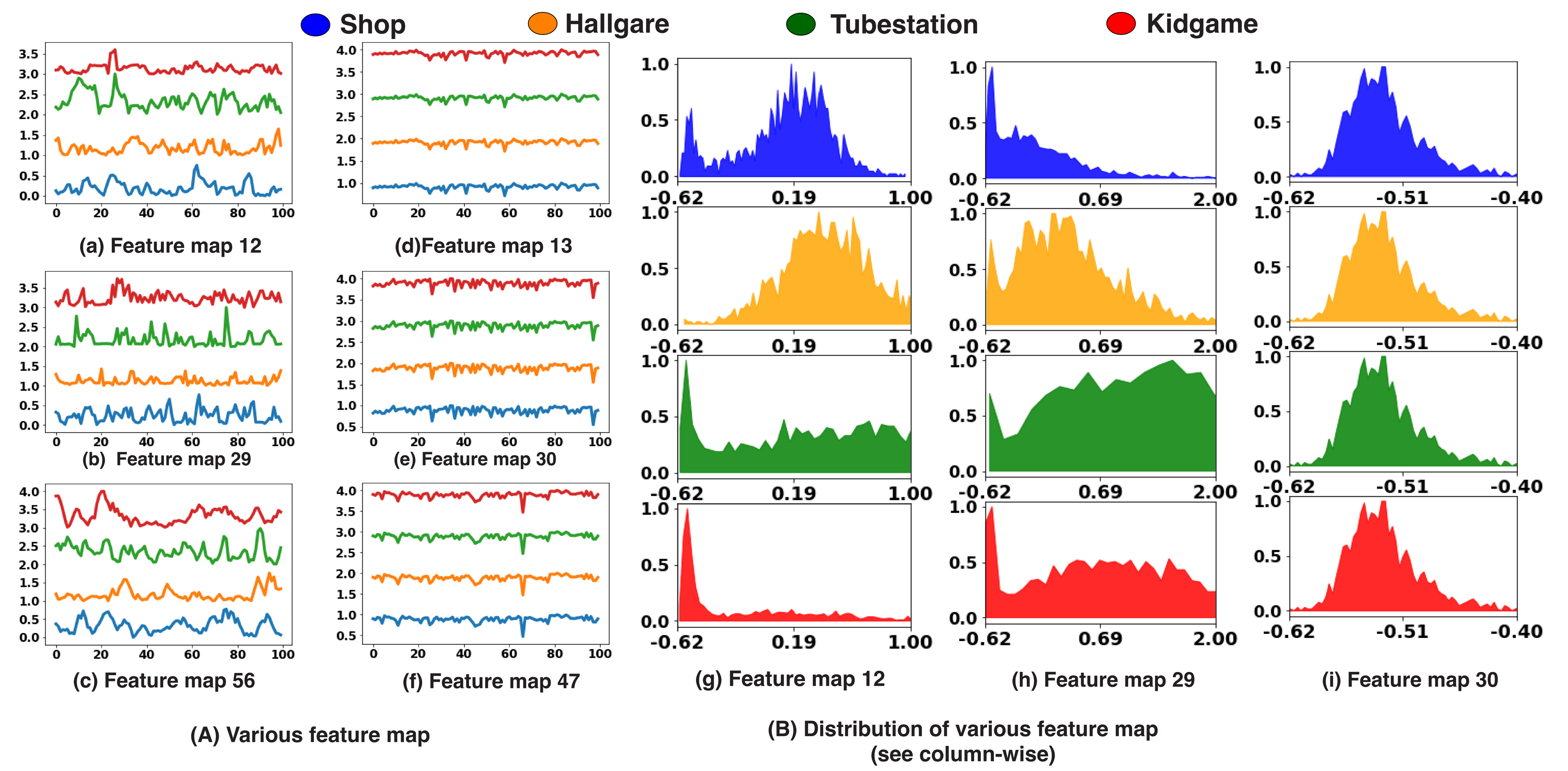
Arshdeep Singh   Padmanabhan Rajan   Arnav Bhavsar
MANAS Lab, School of Computing & Electrical Engg., IIT Mandi, India
d16006@students.iitmandi.ac.in, bopraiarshdeep@gmail.com

ICASSP 2019

Introduction 1

- Despite the success of deep convolution neural networks (CNNs) trained on large-scale dataset, large number of parameters is still a bottleneck since they require more memory, consumes more power in real-time implementation. To optimize the network, redundant parameters in the network can be eliminated.
- We define redundancy of CNNs in terms of feature maps and hypothesis that the redundant feature maps respond similarly to various classes. Hence, participate insignificantly in providing discrimination.
- We employed statistical methods to identify and ignore the feature maps in an ensemble framework as proposed in our previous work [1] on SoundNet [2] for audio scene classification.
- The experiment evaluation on DCASE-16 Evaluation and ESC-50 dataset shows the effectiveness of the proposed approach.

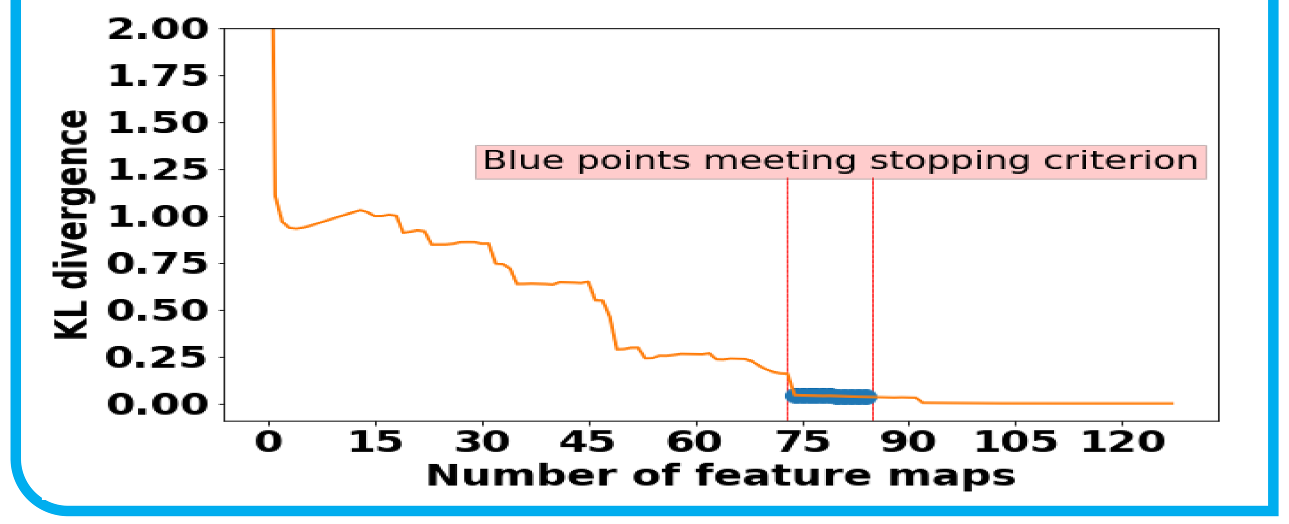
Motivation 2



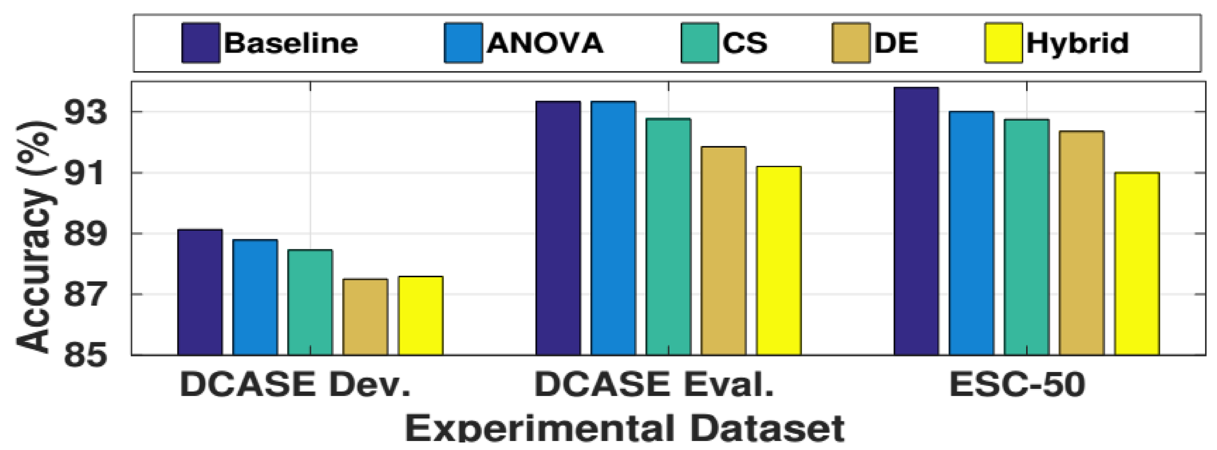
(A) Various feature map

(B) Distribution of various feature map (see column-wise)

KL-Divergence plot 4

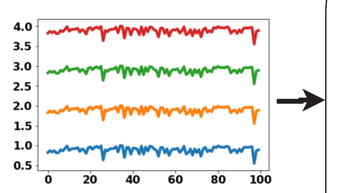


Performance 7

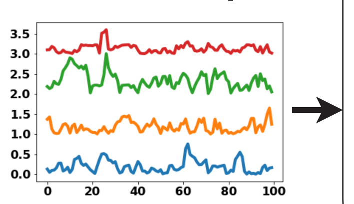


Overall Proposed Pruning Framework 3

Feature map 1



Feature map N



Feature map Ranking

ANOVA method

$$F = \frac{\left(\sum_{n=1}^{p}\sum_{d=1}^{s}(x_{nd}-x_t)^2 - \sum_{n=1}^{p}\sum_{d=1}^{s}(x_{nd}-\bar{x}_n)^2\right)}{\frac{df_b}{df_w}\times\sum_{n=1}^{p}\sum_{d=1}^{s}(x_{nd}-\bar{x}_n)^2}$$

Entropy method

$$\hat{h}_k(\mathbf{x}) = \frac{-1}{p}\sum_{n=1}^{p}\log \hat{p}_k(\mathbf{x_n}),$$

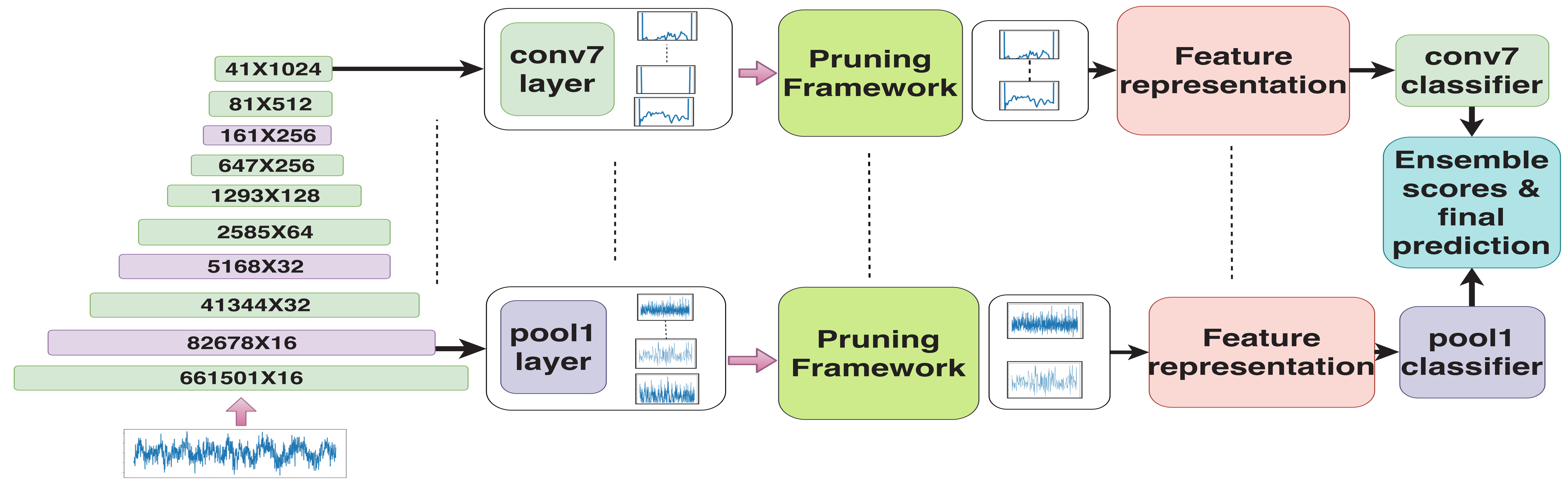
Cosine-similarity method

$$\sigma_z = \sqrt{\frac{1}{p-1}\sum_{n=1}^{p}(\cos\phi_{nz} - \overline{\cos\phi_z})^2}$$

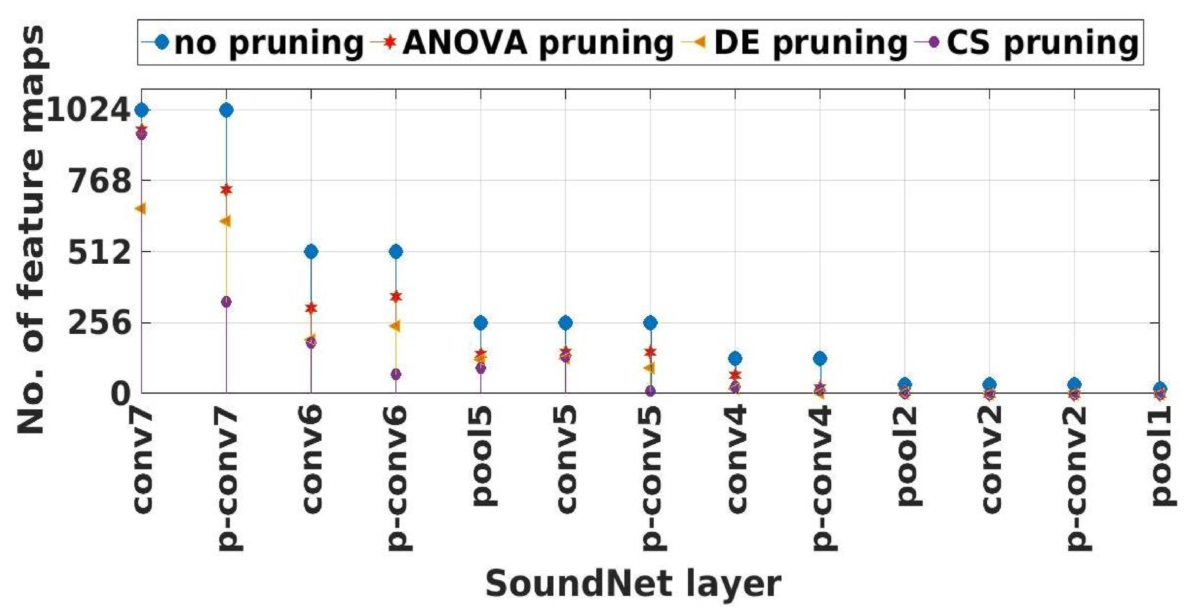
Algorithm 1 A greedy algorithm to select top-l feature maps

```
input  : Λ: Ranked indexes of feature maps.
         H: Distribution of all feature maps of p examples.
         P: A set of N elements each of size R^{p×s}.
         ρ and ϵ are stopping criterion hyper-parameters.
output: Indexes of top-l feature maps (ξ).
# Compute KL div. b/w partial and full set of feature maps.
for m ← 1 to size(Λ) do
    λ = [ ]                      # initialize partial set as an empty set.
    λ.append(Λ[1 : m])           # append first m feature map index.
    h : distribution of feature maps indexed by λ of p examples.
    Ψ[m] = D_KL(h||H)
end
# Select top l feature maps meeting stopping criterion.
for l ← 1 to size(Ψ) do
  # Calculate the change in KL div. b/w each consecutive ρ indexes.
    for w ← 0 to ρ − 1 do
        δ[w] = Ψ(l + w) − Ψ(l + w + 1)
    end
  # Verify the stopping criterion.
    if ||δ||_1 ≤ ϵ then
        ξ = Δ[1 : l]
        break                    # return ξ, the selected indexes.
    end if
end
```

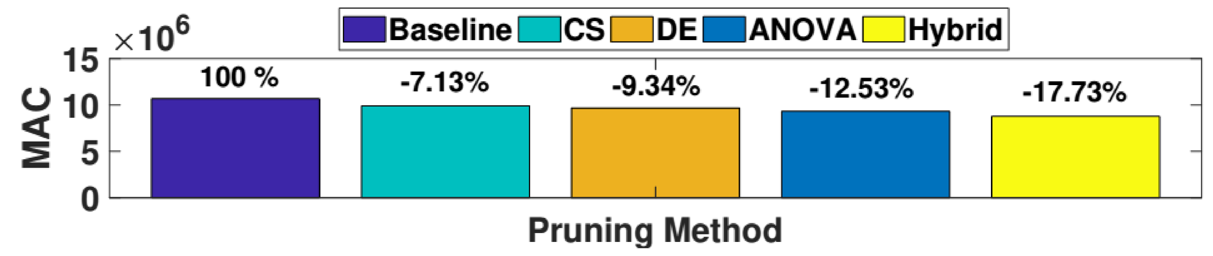
Evaluation framework 5



Layer-wise F.maps 6



MAC operations saved 8



Conclusion

- The statistical methods are being proposed to identify and eliminate redundancy in an ensemble framework.
- The method can be used to prune CNNs based on redundant feature maps only.

References

[1] A. Singh, A. Thakur, P. Rajan and A. Bhavsar, "A Layer-wise Score Level Ensemble Framework for Acoustic Scene Classification," 2018 26th European Signal Processing Conference (EUSIPCO), Rome, 2018, pp. 837-841.
[2] Y. Aytar, C. Vondrick, A. Torralba, "SoundNet: Learning sound representations from unlabeled video" in Advances in Neural Information Processing Systems, pp. 892-900, 2016.