

Deep Pyramidal Residual Networks with Stochastic Depth

大阪府立大学 知能メディア処理研究室
Yoshihiro Yamada, Masakazu Iwamura, Koichi Kise



背景

- 画像から「車両」「猫」といった多様な物体を認識するタスクは**一般物体認識**と呼ばれており世界的企業や研究機関による熾烈な精度争いが進んでいる
- その中で我々は昨年12月のarXiv発表時点で一般物体認識用データセット**CIFAR-10**及び**CIFAR-100**においてその時点の**世界一の認識精度**を達成した

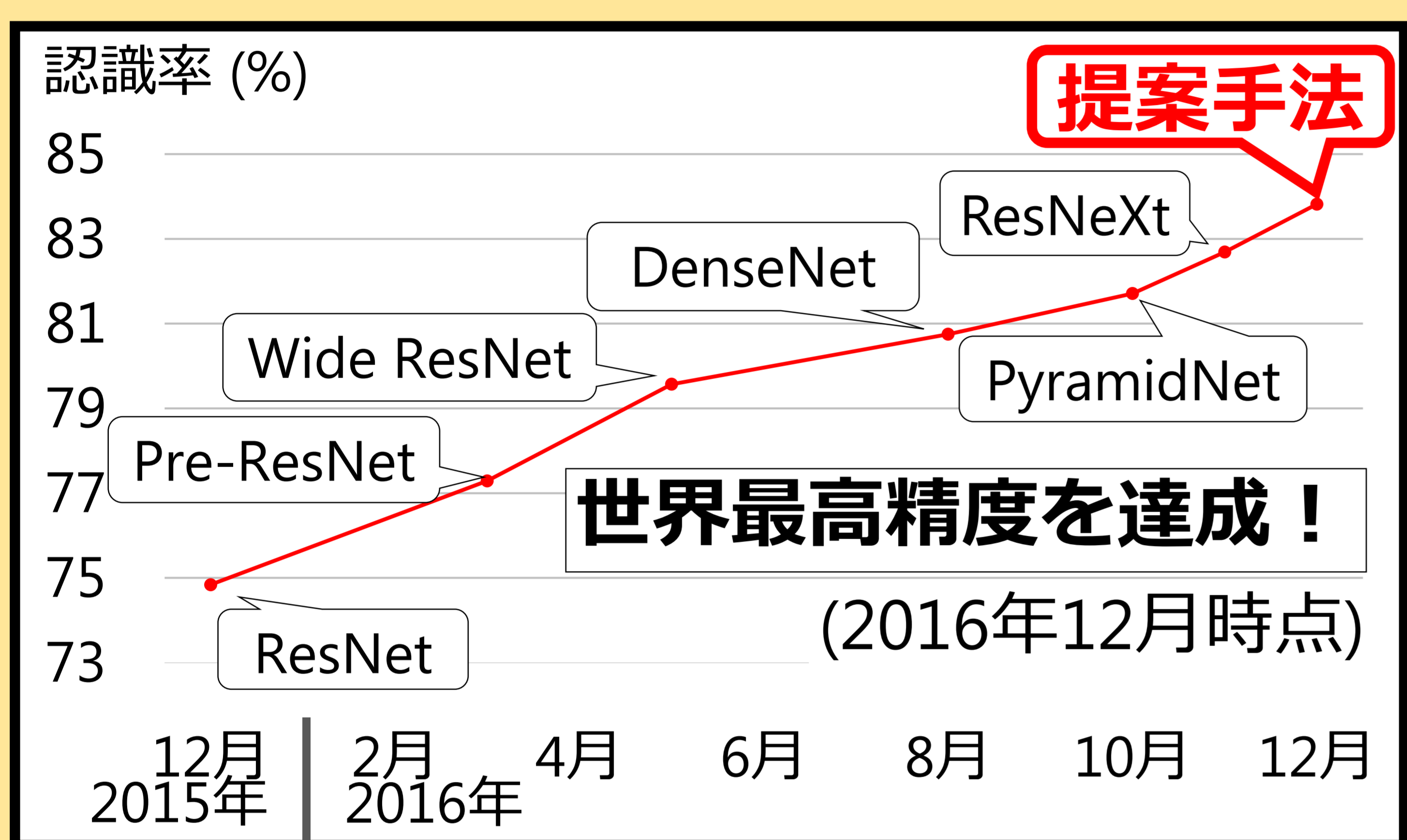


図1. CIFAR-100の最高精度の推移

提案手法

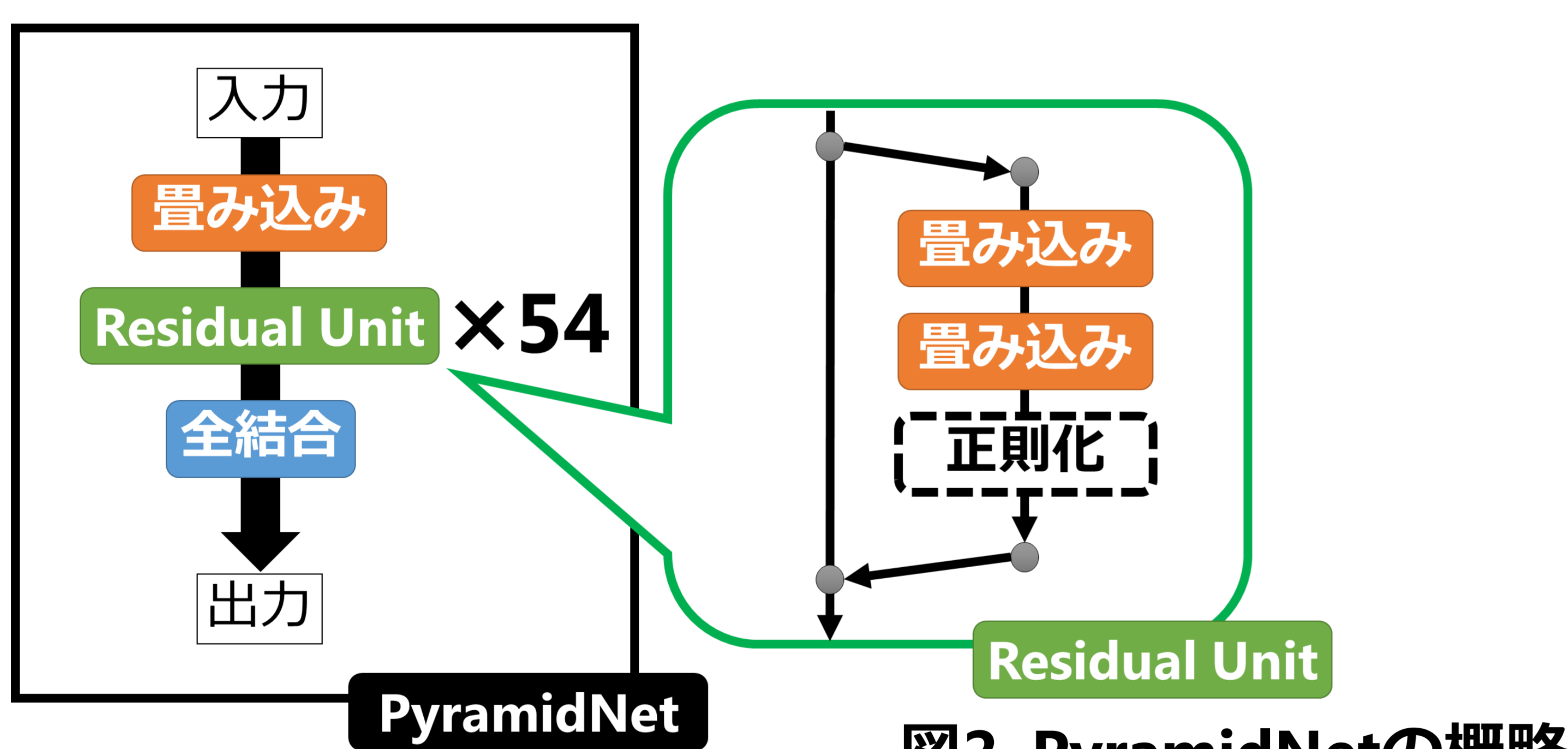


図2. PyramidNetの概略

- PyramidNet (既存手法 [1])**
正則化と呼ばれる
データの間中処理を含まない
- PyramidDrop (提案手法1)**
PyramidNetに対して確率pで出力を0にする正則化を導入
- PyramidSepDrop (提案手法2)**
入力を二つに分割した上でそれぞれについて確率pで出力を0にする正則化を導入

実験(CIFAR-100)

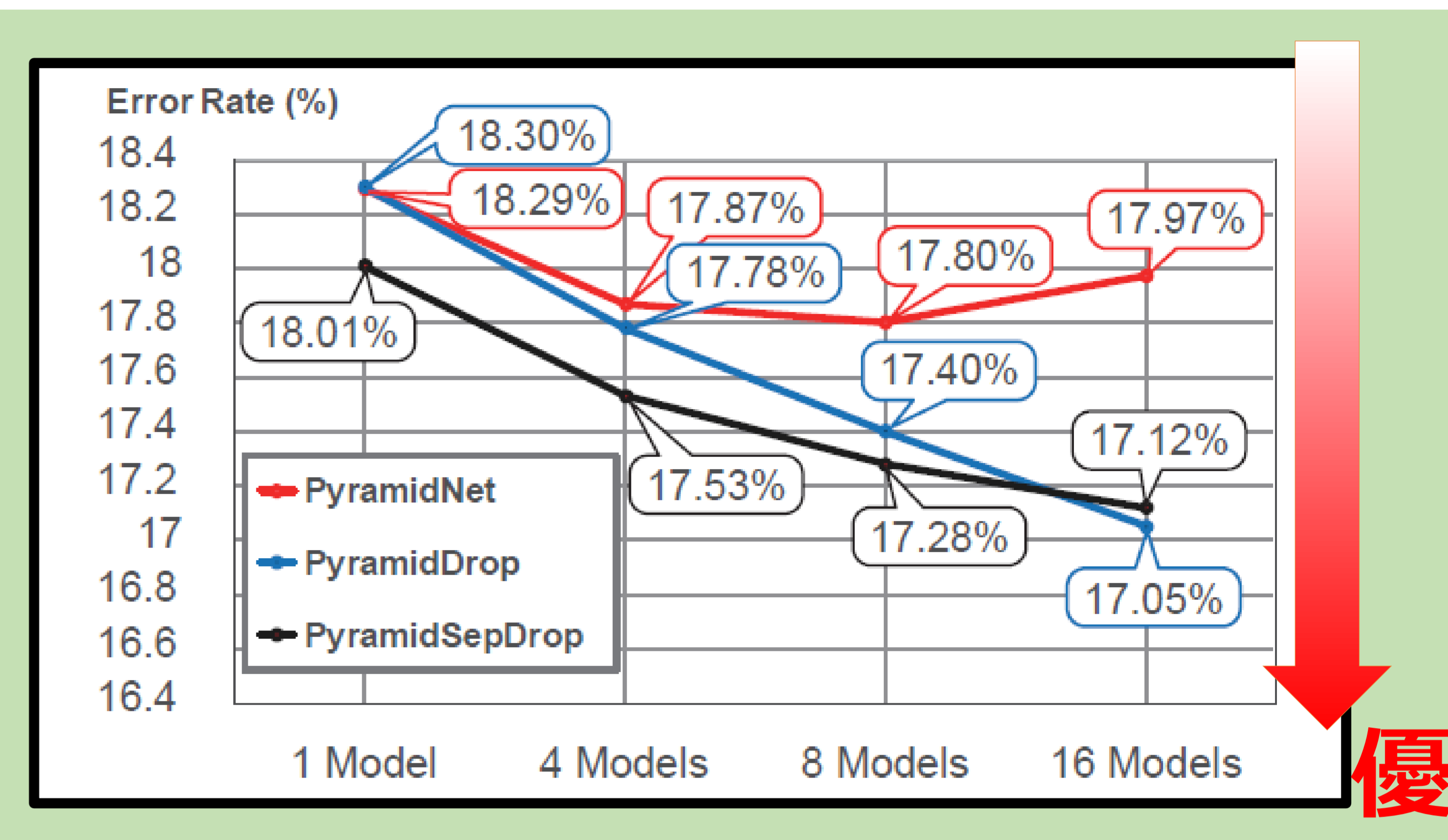


図3. 並列数を増やした場合のError率の比較

- 並列数を増やした場合**PyramidNet**よりも**PyramidDrop**や**PyramidSepDrop**が有効だった
 - 並列数の増加による精度の改善は類似の手法ResNetとResDropの比較で確認できておらず**提案手法独自の傾向**と考えられる
- 層数を増やした場合についても**PyramidNet**よりも**PyramidDrop**や**PyramidSepDrop**が有効だった
- ただし層数と並列数の両方を増やした場合に関して精度こそ及ばなかったものの**PyramidNet**の精度改善が他の2つの手法よりも優れていた

まとめ

- PyramidNetに対する正則化の導入が認識精度にもたらす影響を調べ**並列学習の効果**を明らかにした
- 層数が増えると並列数が少ない場合の認識精度は改善したが並列数が多い場合の精度改善は頭打ちになった
- 検討の中で**世界一の認識精度83.87%**を達成した (今年5月21日にShake-Shakeが発表されるまで)
- なお予稿投稿後7月17日にPyramidNetの改善手法が提案されたことでPyramidDropはパラメータ数を削減と精度改善を達成し**最先端の手法であるShake-Shake[2]とも競争力のある手法**となっている (8月4日時点)

表1. 現時点で精度が上位の3手法とPyramidDropの比較

手法名	#params	認識率
PyramidNet [1]	26.0M	83.65%
ResNeXt [2]	34.4M	83.66%
Shake-Shake [2]	34.4M	84.15%
PyramidDrop	26.0M	84.06%

[1] Dongyoon Han, Jiwhan Kim, and Junmo Kim, "Deep Pyramidal Residual Networks". CVPR 2017. (2017/7/17)

[2] Xavier Gastaldi. "Shake-Shake regularization". arxiv:1705.07485. (2017/5/21)