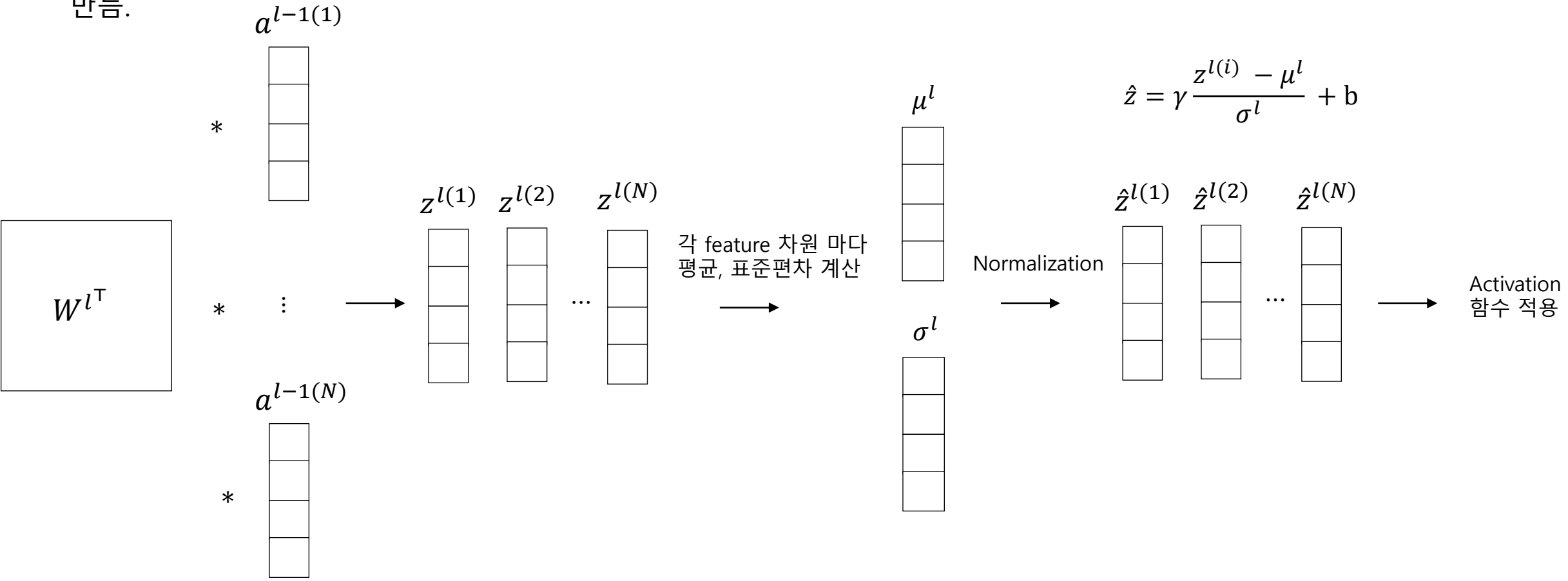


Weight Standardization

- Batch-size에 의존적이고, RNN에는 적용하기 힘든 batch normalization의 단점을 보완하기 위해 제안됨
- Procedure:

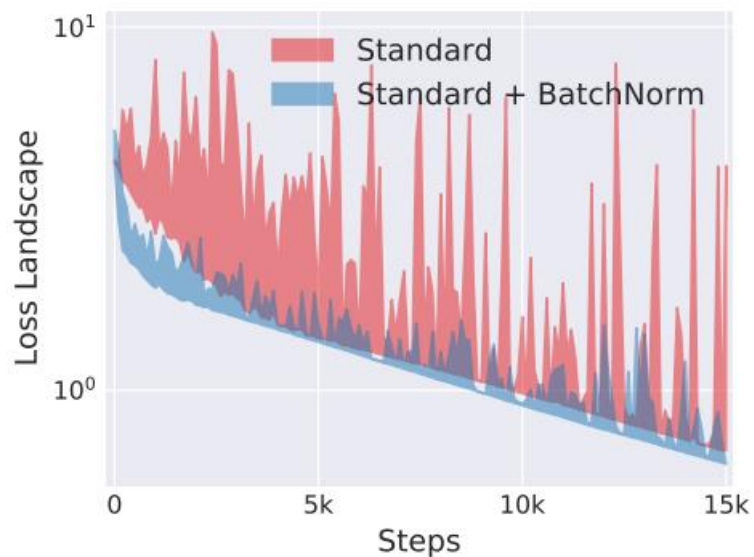
Batch Normalization

- Batch Normalization은 다음 레이어에 입력될 activation 값을(정확히는 activation 함수를 적용하기 전의 summed feature 값) zero-mean, unit variance로 정규화 함으로써 학습 속도를 높이고 loss surface를 부드럽게 만듦.

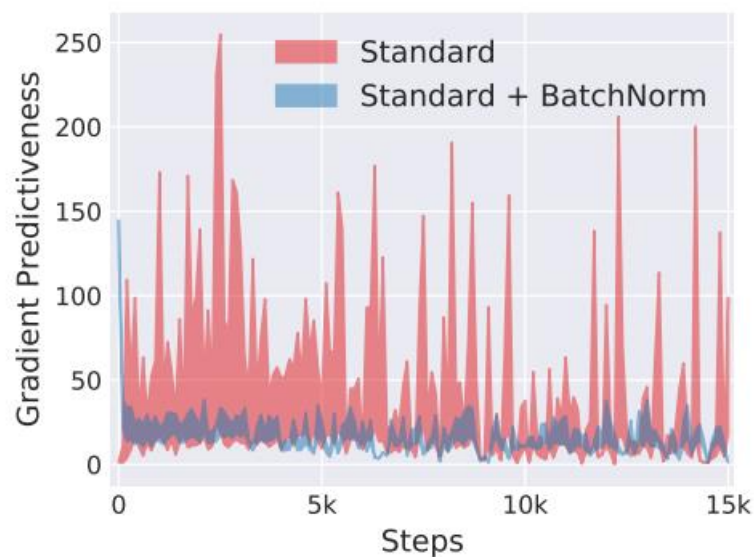


Batch Normalization

- Batch normalization을 적용하였을 경우 loss의 landscape가 훨씬 부드럽게 됨(smoothing 효과).
- Weight에 대한 gradient의 크기(l_2 norm) 또한 batch normalization을 적용하지 않았을 때보다 훨씬 안정됨.



(a) loss landscape

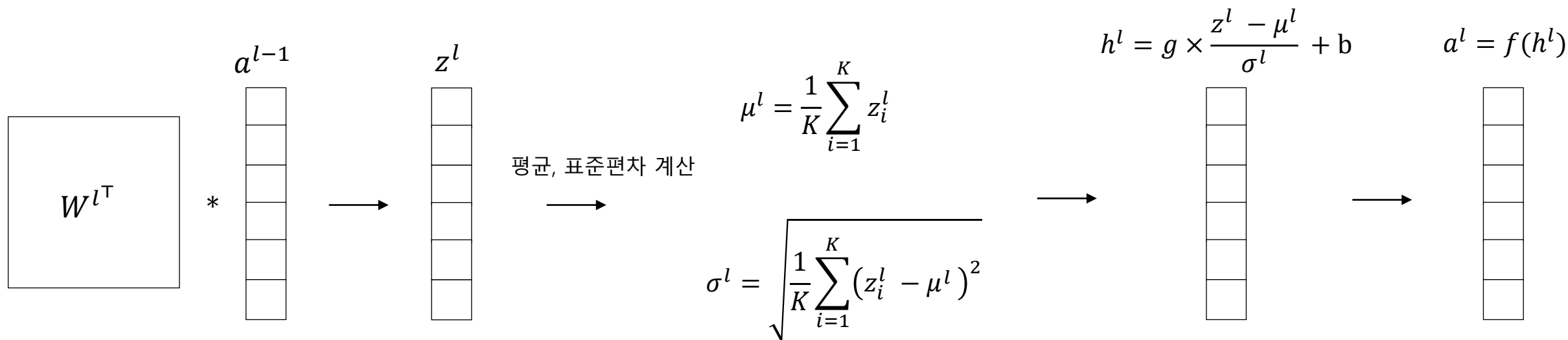


(b) gradient predictiveness

Layer Normalization

- Batch-size에 의존적이고, RNN에는 적용하기 힘든 batch normalization의 단점을 보완하기 위해 제안됨

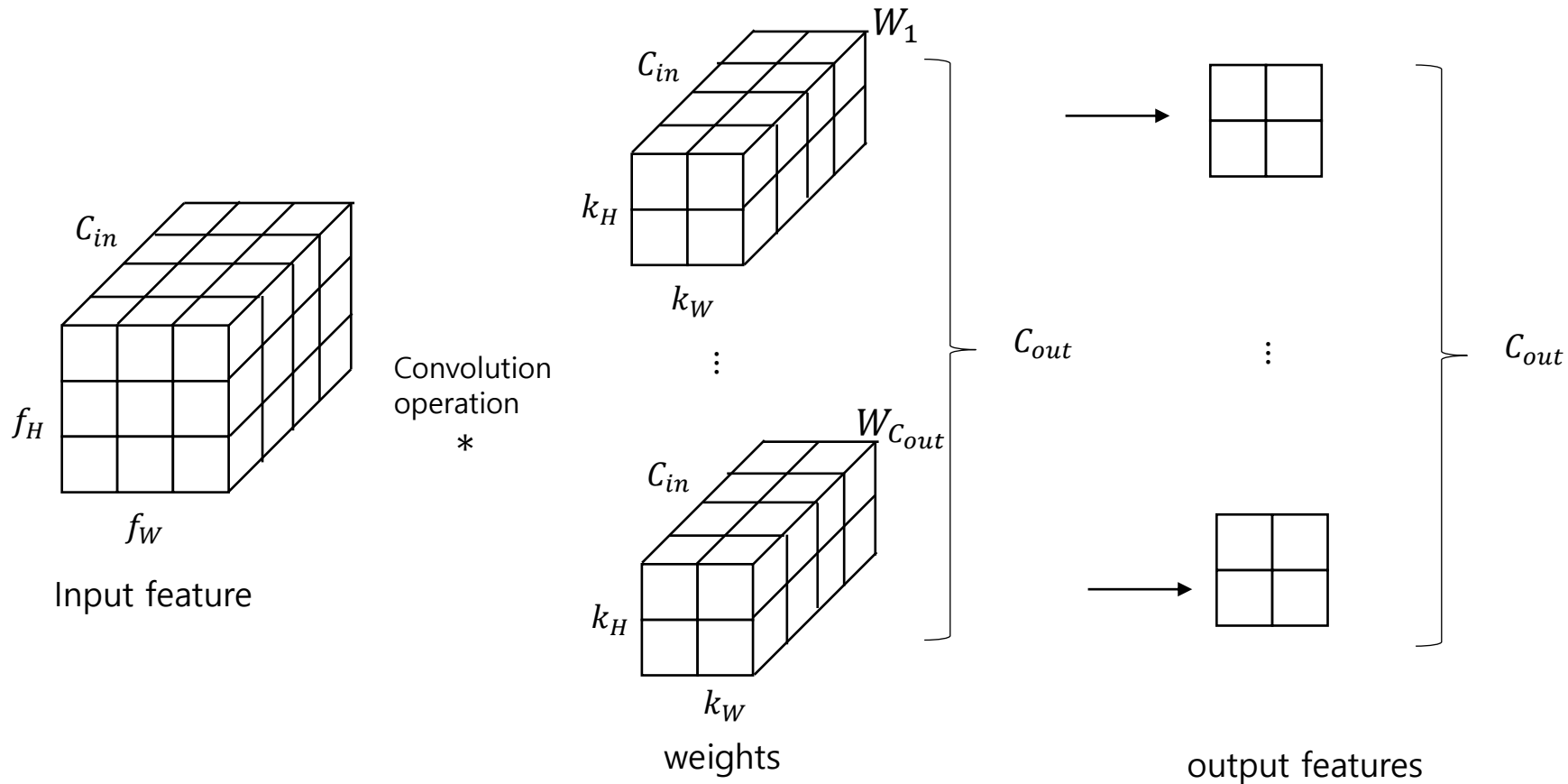
(g : gain parameters(batch-normalization에서 γ 에 해당))



- 각 데이터마다 normalization을 수행하기 때문에 batch-size에 상관이 없음!
- RNN에 적용 가능하지만 CNN에서는 좋은 성능을 내지 못한다.
- Batch Normalization보다 성능에서 우수하지는 않음.

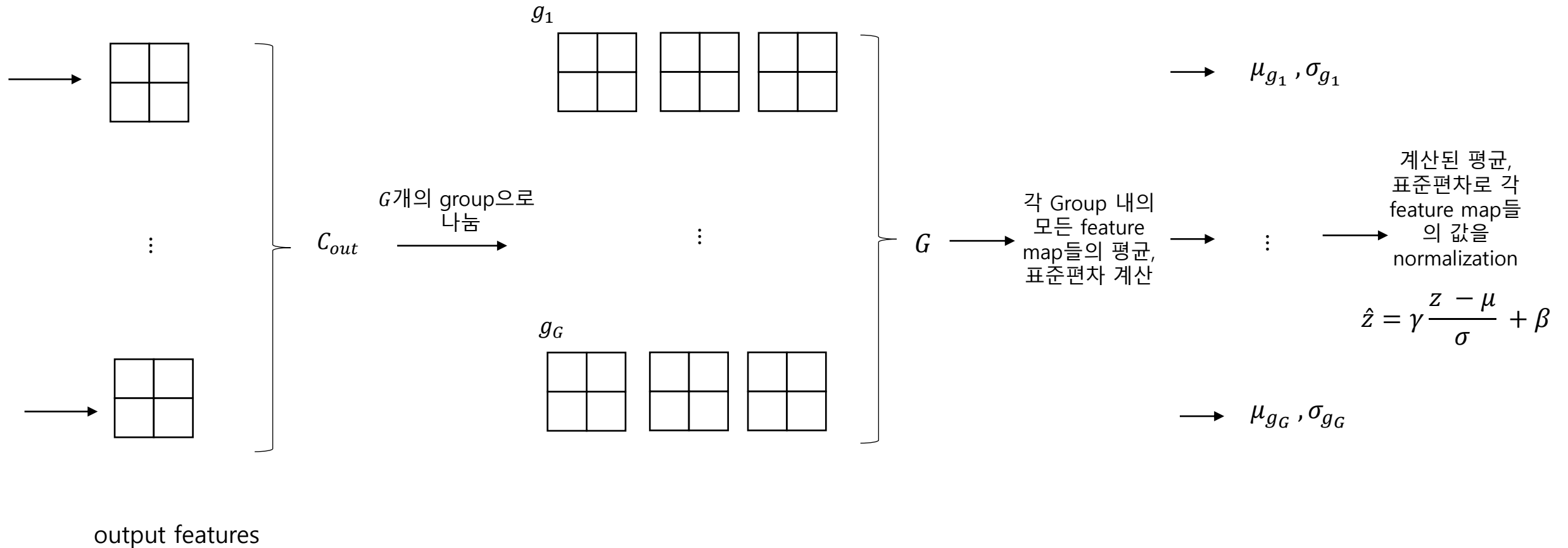
Group Normalization

- Batch-size에 의존적인 Batch normalization의 단점을 보완하기 위해 제안됨.
- CNN에 적용했을 때 성능이 좋지 않은 Layer Normalization을 보완함.



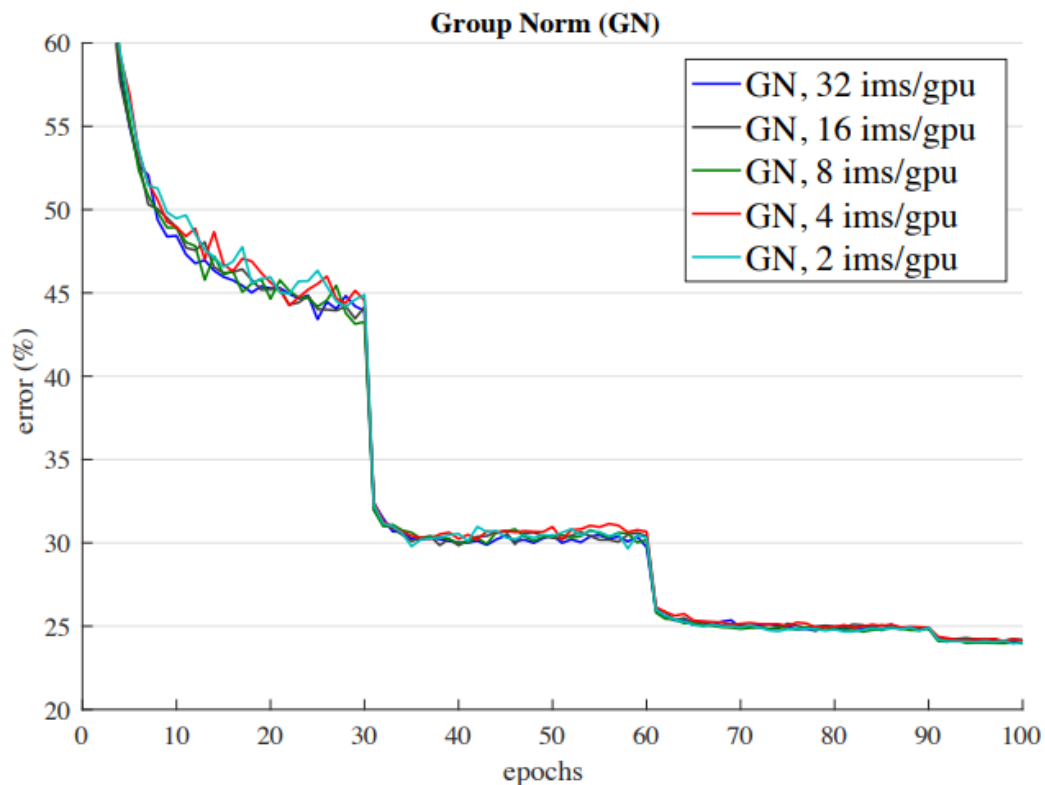
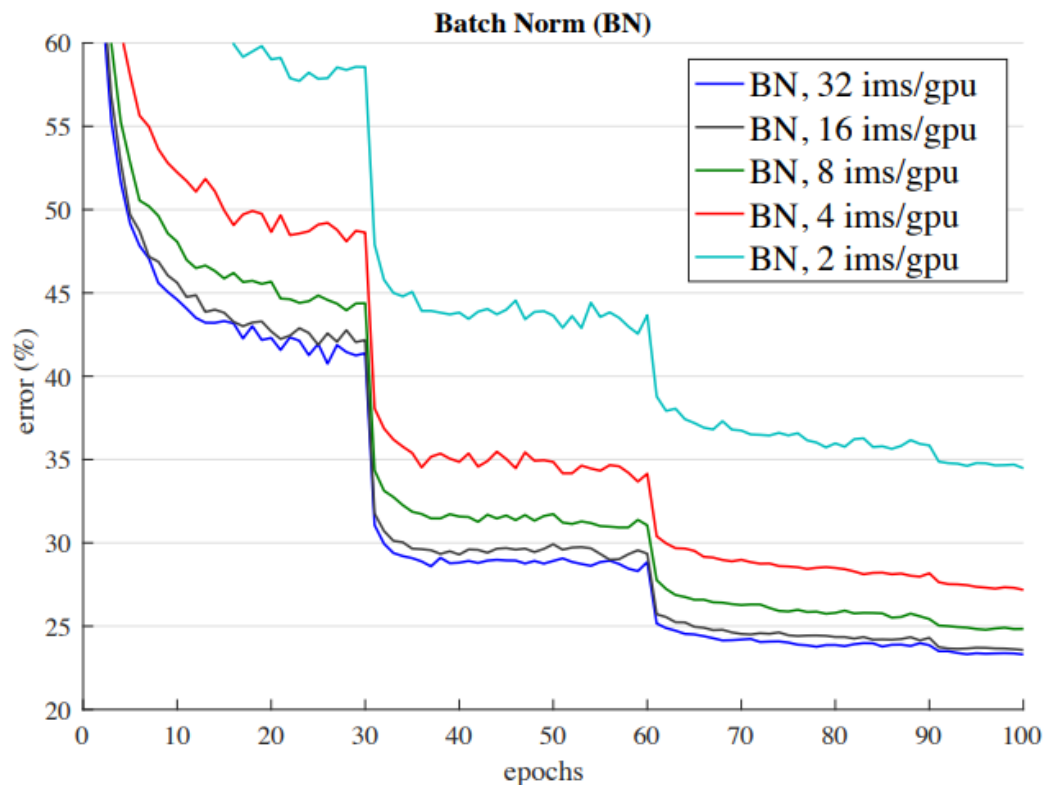
Group Normalization

- Batch-size에 의존적인 Batch normalization의 단점을 보완하기 위해 제안됨.
- CNN에 적용했을 때 성능이 좋지 않은 Layer Normalization을 보완함.



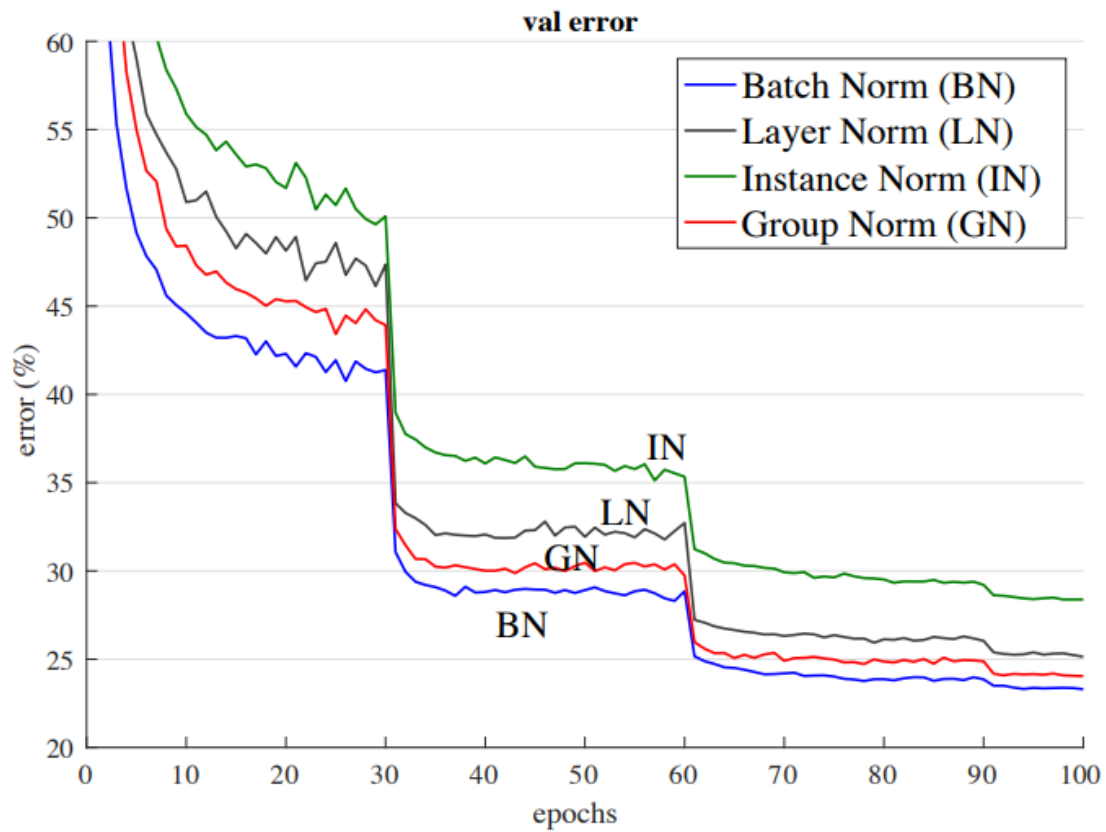
Group Normalization

- Batch normalization은 batch size에 민감하게 반응하지만,
Group normalization은 batch size에 상관없이 동일한 성능을 보임



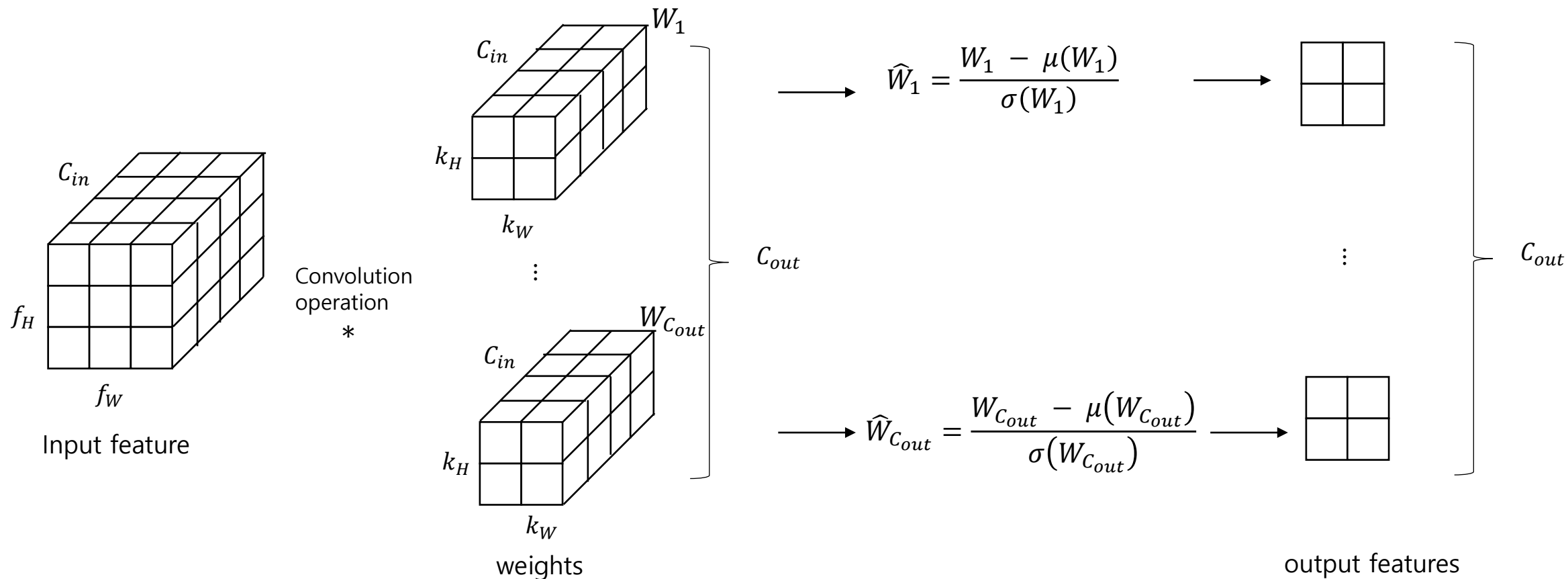
Group Normalization

- 하지만 batch-size의 크기가 충분히 클 때(batch-size = 32), batch normalization의 성능을 넘지는 못함.



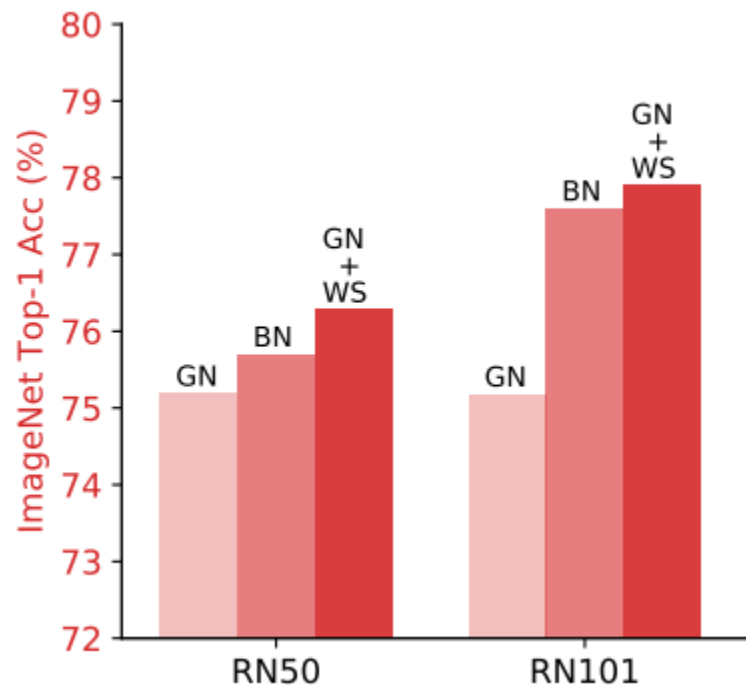
Weight Standardization

- Batch-normalization의 loss landscape를 smoothing 하는 효과를 weight 표준화를 통해 달성하기 위해 제안됨.
- 다른 normalization 기법(batch, layer, group normalization 등)과 결합할 수 있음.



Weight Standardization + Group Normalization

- Weight Standardization 논문에서 제안하는 방법.
- Weight Standardization(WG)과 Group Normalization(GN)을 함께 적용하였을 시 Batch normalization(BN)보다 높은 성능을 달성함.



BN은 큰 batch size로 훈련되었지만, GN과 GN+WS는 batch size = 1로 훈련되었음.