Signal Parameter Estimation Using Complex-Valued Deep Reinforcement Learning

Yuting Ng, Chin Yuan Chong
SONDRA, CentraleSupélec

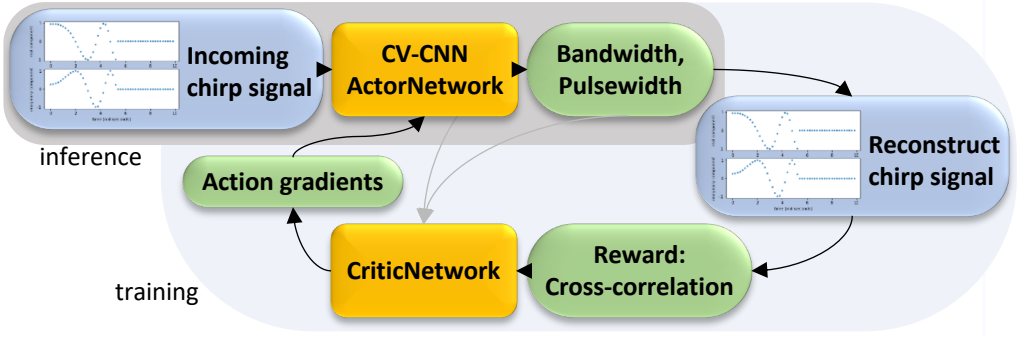
1. Objectives

- Estimate bandwidth and pulsewidth of chirp, a signal which frequency sweeps (increases/decreases) with time
- Complex-sampled signal as input; unsupervised training approach that incorporates signal processing techniques

2. Approach

- Complex-valued convolutional neural network (CV-CNN) as first layer of neural network
- Deep deterministic policy gradient (DDPG) for training

3. Pipeline



4.Complex-Valued Convolutional Neural Network (CV-CNN)

- Complex first layer has two filters for every feature map [1,2]
- Nonlinearity $\mathrm{f(A,B)} = \sqrt{A^2 + B^2}$ gives real values at output

5. Deep Reinforcement Learning (RL): Deep Deterministic Policy Gradient (DDPG)

- Classical RL techniques have discrete state and action spaces
- Deep RL techniques able to handle continuous spaces:

Techniques	Continuous		Application
	State	Action	
Deep Q Network (DQN) [3]	✓		Atari games
Deep Deterministic Policy Gradient (DDPG) [4]	✓	✓	Physics tasks

- DDPG: Environment, ActorNetwork and CriticNetwork

6. Signal Processing as Environment

- Generate chirp from bandwidth and pulsewidth
  bw : U(0,1) MHz
  pw : U(0,10) msec
  Φ : U(0,1) if incoming
  Fs : 5 MHz

$$s[t] = \begin{cases} e^{-j\pi(\frac{bw}{pw}t^2+\phi)}, & t \le pw \\ 0, & t > pw \end{cases}$$

- Cross-correlation as reward

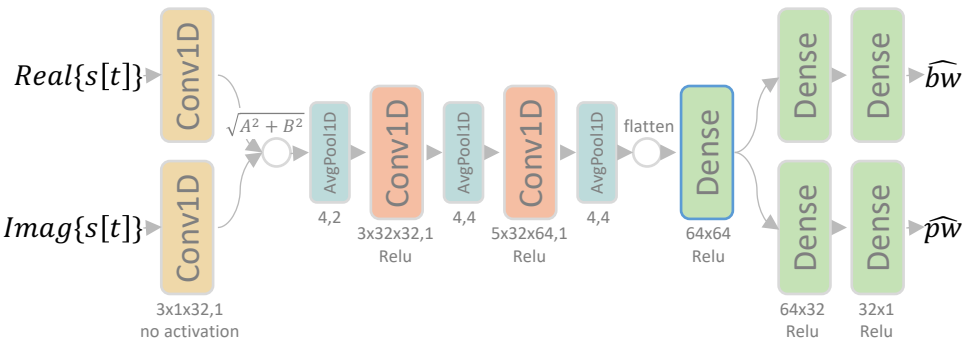
$$r = \|s[t]^H \hat{s}[t]\|$$

$$r = -0.1, r < 0.5$$

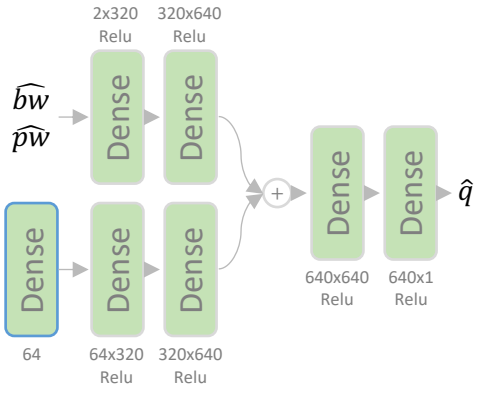
References

[1] Tygert et al, "A mathematical motivation for complex-valued convolutional networks", *Neural Computation*, vol. 28, no. 5, pp. 815-825, 2016.
[2] Wilmanski et al, "Complex input convolutional neural networks for wide angle SAR ATR", in 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Dec 2016.
[3] Mnih et al, "Human-level control through deep reinforcement learning", *Nature*, vol. 518, pp. 529-533, 2015.
[4] Lillicrap et al, "Continuous control with deep reinforcement learning", in 2016 International Conference on Learning Representations (ICLR), May 2016.

7. ActorNetwork with CV-CNN



8. CriticNetwork and Training



Train CriticNetwork

- $q = r + \gamma\hat{q}$
- $\text{loss} = (q - \hat{q})^2$

CriticNetwork output

- grads: $\frac{\partial q}{\partial bw}, \frac{\partial q}{\partial pw}$

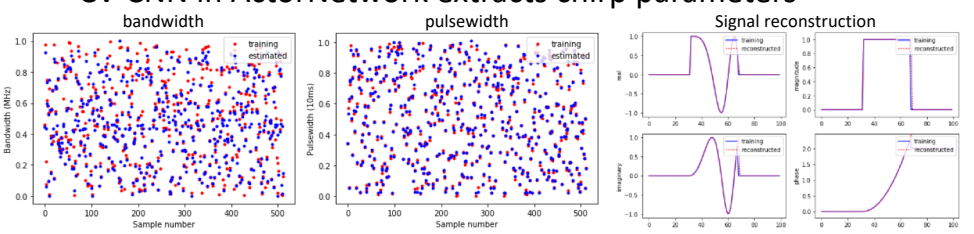
Train ActorNetwork with

- grads: $\frac{\partial q}{\partial AN} = \frac{\partial q}{\partial bw}\frac{\partial bw}{\partial AN} + \frac{\partial q}{\partial pw}\frac{\partial pw}{\partial AN}$

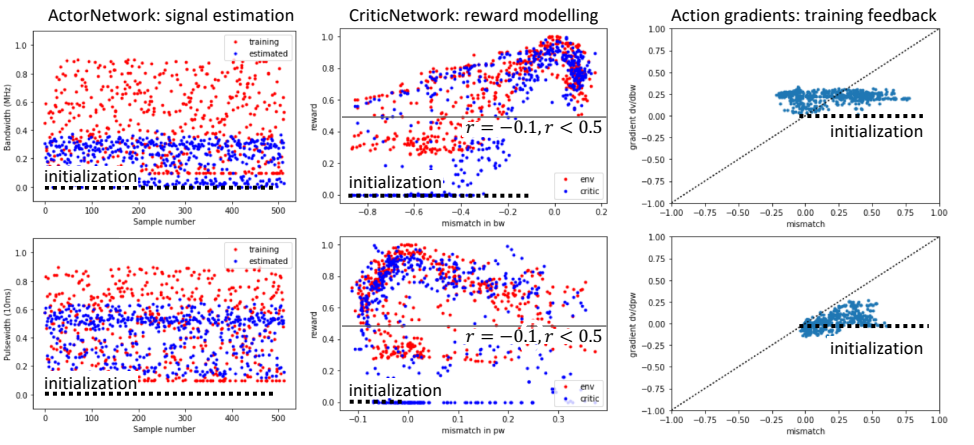
Batch size = 512; Target networks with $\tau = 0.001$; Action exploration noise = OU($\theta = 0.6, \sigma = 0.01$);
Learning rate CriticNetwork = 0.0001; $\gamma = 0.99$; Learning rate ActorNetwork = 0.00001; ϵ −= 0.001.

9. Results and Discussion

- CV-CNN in ActorNetwork extracts chirp parameters



- CriticNetwork models correlation reward
- Correlation has multiple "sidelobes" as local minimas, use threshold of 0.5 to remove local minimas during training
- Training snapshot, (dotted black line indicates initial output):



10. Conclusion

- Proposed unsupervised approach incorporating signal processing techniques for parameter estimation of complex-valued signals.
- Applied approach to bandwidth and pulsewidth estimation of chirp signals.
- More investigation required for crafting of reward function and training using DDPG algorithm.