Achieving Computational Robustness

Adaptive-time computation model

computing models (Image2Text, Seq2Seq)

Q: Given a model $M$, a sample input $x$, does the model terminates in $K$ steps for all inputs $x’$ close to $x$?

\[
L(M(x)) = \max_{i \leq n} \left\{ \sum_{T_i} \right\}
\]

\[
\text{Input Space: } X \quad \text{Perturbation Space: } \delta \in \mathbb{R}^n
\]

\[
\text{Verifiable: Constraint Encoding}
\]

\[
\text{Testing: Adversarial Attacks}
\]

\[
\text{Prov: max } L(M(x)) \leq K
\]

\[
\text{Verification: Constraint Encoding}
\]

\[
\text{Completo but Expensive}
\]

\[
\text{Scalable but No Guarantee}
\]

\[
\text{Problem}
\]

\[
\text{Robustness Objective}
\]

\[
\text{Our Approach}
\]

\[
\text{Motivation}
\]