Fine-grained Categorization and Dataset Bootstrapping using Deep Metric Learning with Humans in the Loop

Yin Cui\textsuperscript{1,2}  Feng Zhou\textsuperscript{3}  Yuanqing Lin\textsuperscript{3}  Serge Belongie\textsuperscript{1,2}

\textsuperscript{1}Department of Computer Science, Cornell University \textsuperscript{2}Cornell Tech \textsuperscript{3}NEC Labs America

Motivation

FGVC Challenges

➢ Lack of training data.
➢ Large number of categories.
➢ High intra-class vs. low inter-class variances.

Proposed Solutions

qd Bootstrapping training data from the web.
qd Learning compact low-dim representations.
qd Learning manifolds with multiple anchor points.

Framework

Softmax vs. Metric Learning

\begin{align*}
\text{CNN with softmax loss} & \quad \text{CNN for metric learning} \\
\end{align*}

\begin{itemize}
\item Pre-defined one-hot encoding versus learned manifold.
\item Compared with Softmax, metric learning could learn a more compact representation in a much lower dimensional space.
\end{itemize}

Learning Manifolds

Hard Negatives

\begin{itemize}
\item \(\Omega(n^3)\) possible triplets, impossible to go through.
\item Training from hard negatives by:
  \begin{enumerate}
  \item Only keeping triplets that violate constraint.
  \item Including human-labeled false positives.
  \end{enumerate}
\end{itemize}

Local Positives

\begin{itemize}
\item Sampling local positives could learn a more spread manifold rather than a dense sphere.
\end{itemize}

Learning Anchor Points

\begin{itemize}
\item Incorporating class labels into metric learning.
\item Back-propagate classification loss to update anchor points.
\end{itemize}

Contributions

\begin{itemize}
\item A unified framework for simultaneous fine-grained categorization and dataset bootstrapping.
\item A novel metric learning method that learns manifolds from both machine-mined and human-labeled hard negatives.
\item A fine-grained flower dataset with 620 categories and around 30K images.
\end{itemize}

Experiments

Original Flower-620 (15K images)

\begin{itemize}
\item Softmax (68.8)
\item Triplet-A (68.5)
\item Triplet-M (66.8)
\item Triplet-H (66.3)
\item Triplet-I (65.9)
\item Triplet-M (64.4)
\item Softmax + HNS (80.1)
\end{itemize}

\begin{itemize}
\item Softmax + HNS (68.0)
\item Softmax + HNS (80.1)
\item Softmax + HNS (86.6)
\item Softmax + HNS (70.3)
\item Softmax + HNS (70.2)
\item Softmax + HNS (70.8)
\end{itemize}

\textsuperscript{[5]} Metric Learning: +2.7\% over softmax, with a much more compact representation.

\textsuperscript{[6]} Dataset Bootstrapping: +6.9\% (+3.4\% from new data, 3.5\% from human-labeled hard negatives).

Visualization of flower embedding

\begin{itemize}
\item Incorporating class labels into metric learning.
\item Back-propagate classification loss to update anchor points.
\end{itemize}

\begin{align*}
\mathcal{L}_{\text{triplet}}(x, x_p, x_n) &= \max \left\{ 0, \| f(x) - f(x_p) \|^2 - \| f(x) - f(x_n) \|^2 + m \right\}
\end{align*}