1. **Motivation**
   1. Speicherbedarf: CNN haben Millionen Parameter, Geräte eingeschränkter Speicherplatz.
   2. Große Anzahl von Speicherzugriffen und arithmetischen Operationen->Strom Verbrauch
2. **Ziel der Arbeit.**
   1. Entwicklung einer CNN, das
      1. Zu wenig Parameter braucht.
      2. Gute Inferenzzeit.
      3. Ergebnisse auf den Stand der Technik liefern.
3. **Vergleich zwischen berühmten CNN**
   1. AlexNet
   2. Xception
   3. MobileNet
   4. TemkiNet
   5. Was feht noch in anderen CNN, was kann verbessert werden
      1. Beschreibung.
      2. Vorteile.
         1. Zu wenige Parameter erfordern
         2. Kurze Inferenzzeit
         3. Einfache Implementierung
            1. Hyperparameter: Anzahl der Neuronen pro Schicht.
      3. Nachteile.
         1. Lange Trainingszeit.
4. **Techniken zur Verbesserung der Netzwerkleistung.**
   1. Data Augmentation.
      1. Verdoppelt die originale Netzwerkleistung.
   2. Optimierer.
      1. Adam bessere Ergebnisse
      2. Optimierer hängen sehr von der Lernrate ab
   3. Batch-Normalisierung.
      1. Beeinträchtigung der Netzwerkleistung bei Verwendung einer zu kleinen Lernrate.
      2. Sehr gut bei Verwendung größer Lernrate.
   4. Bildgröße.
      1. Höhere Leistung bei größeren Bildern.
      2. Zu lange Trainingszeit bei größeren Bildern.
   5. Anzahl des Neurons pro Schicht.
      1. Verbesserung der Netzwerkleistung bis eine bestimmte Anzahl mehr und mehr Neuronen pro Schicht.
   6. Aktivierungsfunktion.
      1. Tanh->ReLU6->PreLU->LeakyReLU
   7. Qualität des Datensatzes.
      1. Bessere Leistung mit guten und sorgfältigen ausgewählten Datensatz.
   8. Lernrate.
      1. Kleiner: zu langsame Konvergenz
      2. Größer: keine Konvergenz oder zu viele Schwankungen.
   9. Dropout.
      1. Robuster und stabiler
      2. Kleine Verbesserung der Netzwerkleistung
5. **Reduktion der Netzwerkgröße und Erhöhung der Inferenzzeit.**
   1. **Extreme Versionen.**
      1. Weniger Parameter -> Weniger Speicherplatz.
      2. Bessere Inferenzzeit.
      3. Gleiche Genauigkeit oder vernachlässige Verlust der Genauigkeit im Vergleich zu originalem Modell.
   2. **Pruning**
      1. Sparsamkeit im CNN: Reduktion der nicht-nulle Parameter.
      2. CNNs haben überflüssige Parameter.
      3. Kleiner Verlust der Genauigkeit bei großen CNNs.
      4. Großer Verlust der Genauigkeit bei kleinen CNNs.
   3. **Quantisierung** 
      1. 16 Bits
         1. 2x schneller
         2. Weniger Speicherplatzbedarf
      2. 8 Bits
         1. 4x schneller
         2. Weniger Speicherplatzbedarf
6. **Extra**
   1. Neue Lernrate Scheduler
   2. Neue schnelle Trainingstechnik.
      1. Curriculum Lernen.
      2. Destillation
   3. Schnelle Gewicht Initialisierung