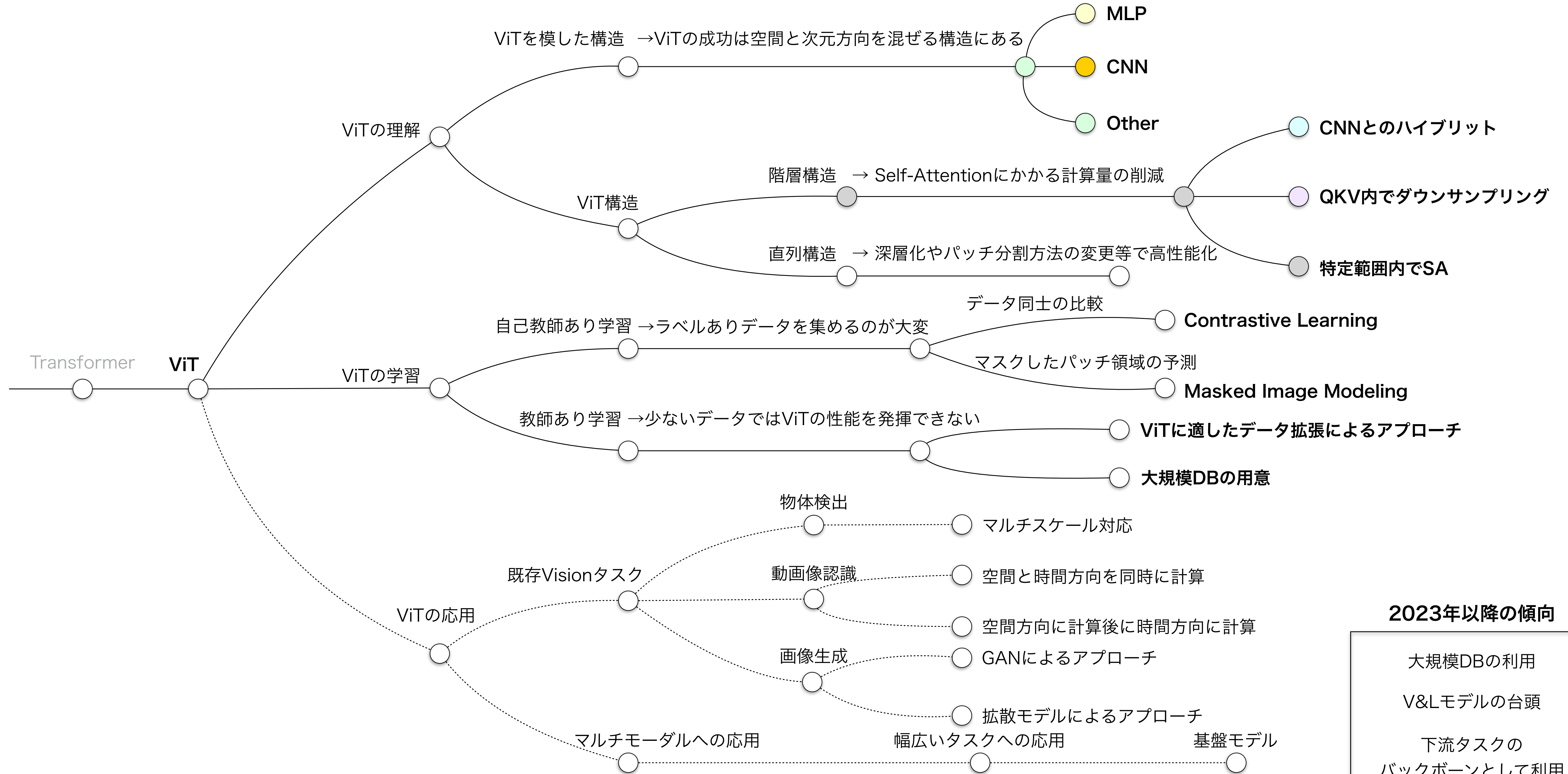


Vision Transformer概要



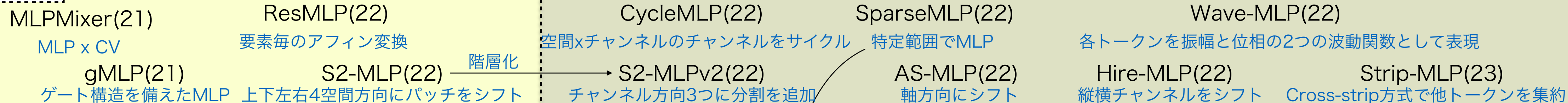
2023年以降の傾向

- 大規模DBの利用
- V&Lモデルの台頭
- 下流タスクのバックボーンとして利用

ViTの構造

ViTを模した派生構造

MLP



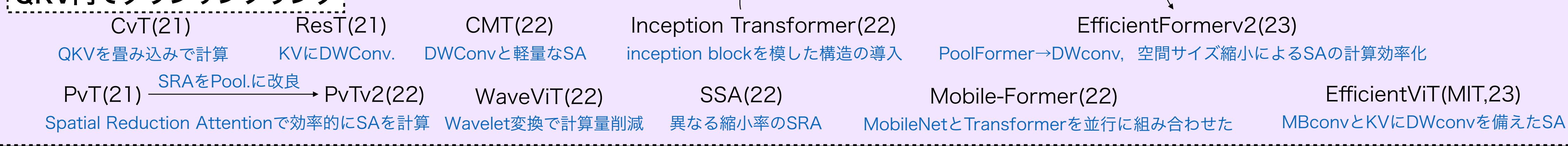
CNN



CNNとのハイブリッド



QKV内でダウンサンプリング



階層構造



ViTの学習

