**Expanded Deep Learning Notes**

**1. Filters, Strides, Padding, Pooling, CNN Architecture, BackPropagation**

**Filters (Kernels)**  
*Definition:* Small matrices applied over input images to extract localized features such as edges, corners, or patterns.  
*When to use:* Feature extraction in convolutional layers, initial stages of image processing.  
*Detailed Explanation:* Filters learn to detect common patterns, like vertical edges or textures, in early layers. Deeper layers extract more abstract concepts. Training adapts each filter, letting the network specialize.  
*Case Study Examples:* Edge detection in medical images; object patterns in self-driving car cameras.

**Strides**  
*Definition:* Step size by which the filter moves across the image.  
*When to use:* To control the output feature map size and reduce computational cost.  
*Detailed Explanation:* Larger strides result in more downsampling but may omit subtle details. Smaller strides preserve more information but are computationally intensive.  
*Case Study Examples:* Fast face detection for mobile devices; quick image scans in robotics.

**Padding**  
*Definition:* Adding extra pixels (usually zeros) to image borders before convolution.  
*When to use:* To preserve output dimensions or prevent edge information loss.  
*Detailed Explanation:* 'Same' padding keeps input size constant, helpful in deep networks. 'Valid' padding shrinks the output, making computations faster.  
*Case Study Examples:* Accurate border detection in satellite imagery.

**Pooling**  
*Definition:* Operations (max/average) that downsample feature maps while retaining essential information.  
*When to use:* To summarize features, reduce computations, and overfitting.  
*Detailed Explanation:* Max pooling keeps the strongest signal within a region, while average pooling smooths the feature map.  
*Case Study Examples:* Feature reduction in VGGNet; summarizing retinal scan features.

**CNN Architecture**  
*Definition:* Models that stack convolution, pooling, and fully connected layers for image analysis.  
*When to use:* Any image, video, and vision-based learning tasks.  
*Detailed Explanation:* Multiple layers extract a hierarchy of features. Techniques like skip connections, normalization, and regularization (dropout) improve training.  
*Case Study Examples:* VGG for categorizing images; ResNet for advanced facial recognition.

**BackPropagation**  
*Definition:* Technique to train neural networks by gradient-based parameter updates.  
*When to use:* Optimizing deep models for lowest error/loss.  
*Detailed Explanation:* Chain rule applied to compute gradients of error with respect to each weight; allows rapid adaptation in deep and wide networks.  
*Case Study Examples:* Training visual classifiers; teaching RNNs to predict text.

**2. VGG, ResNet, EfficientNet**

**VGG**  
*Definition:* CNN using consecutive 3x3 convolutions and max pooling.  
*When to use:* Image classification and feature extraction for transfer learning.  
*Detailed Explanation:* VGG offers a streamlined, modular approach but can be large and slow for deployment.  
*Case Study Examples:* ImageNet competitions; transfer learning for custom datasets.

**ResNet**  
*Definition:* Deep CNN with residual connections (skip links) to alleviate vanishing gradients.  
*When to use:* Building very deep models for challenging recognition tasks.  
*Detailed Explanation:* Residual skips let signals move forward and gradients backward easily, enabling 100+ layer models.  
*Case Study Examples:* Champions in major vision challenges; backbone for detection like Faster R-CNN.

**EfficientNet**  
*Definition:* Models that systematically scale depth, width, and image resolution.  
*When to use:* When speed and accuracy are needed at a low compute cost.  
*Detailed Explanation:* Compound scaling ensures a balanced increase in network power without unnecessary parameters.  
*Case Study Examples:* Mobile image recognition; edge AI cameras.

**3. Object Detection or Image Classification**

**Object Detection**  
*Definition:* Identifying and localizing objects in images using bounding boxes and labels.  
*When to use:* Multiple objects per image or position-aware applications.  
*Detailed Explanation:* Models like YOLO use single-shot detectors for real-time performance; Faster R-CNN uses two-stage detection for accuracy.  
*Case Study Examples:* Self-driving car sensors; surveillance footage analytics.

**Image Classification**  
*Definition:* Assigning a single class to an entire image.  
*When to use:* Single-object or broad category recognition.  
*Detailed Explanation:* Networks like VGG or EfficientNet output one label; often used for tagging or sorting.  
*Case Study Examples:* Dog-vs-cat classifiers; digit recognizers for postal sorting.

**4. RNN, LSTM, GRU Basics**

**RNN**  
*Definition:* Neural networks for sequential data with internal memory.  
*When to use:* Time series, text, and sequential event modeling.  
*Detailed Explanation:* Input is processed one step at a time, with each output dependent on the previous step.  
*Case Study Examples:* Speech recognition; weather forecasting.

**LSTM**  
*Definition:* RNN variant designed to prevent vanishing gradients.  
*When to use:* Long text or time-series sequences with dependencies.  
*Detailed Explanation:* LSTM uses gates to control memory updates, allowing it to remember important information for many steps.  
*Case Study Examples:* Machine translation; text generation.

**GRU**  
*Definition:* A simpler, faster alternative to LSTMs.  
*When to use:* Faster training with similar performance for moderate dependencies.  
*Detailed Explanation:* Combines memory and update gates into simpler structures, requiring fewer parameters.  
*Case Study Examples:* Chatbot dialog prediction; real-time speech understanding.

**5. Text Cleaning, Tokenization, Stop Words**

**Text Cleaning**  
*Definition:* Removing noise, punctuation, or formatting from raw text.  
*When to use:* Before all NLP pre-processing or modeling.  
*Detailed Explanation:* Includes steps like lower-casing, removing unwanted characters, and normalizing text for better downstream model accuracy.  
*Case Study Examples:* Scraping emails for spam filtering.

**Tokenization**  
*Definition:* Splitting text into units (words, subwords, sentences).  
*When to use:* Converting raw text for modeling as numerical sequences.  
*Detailed Explanation:* Can be at word, character, or sentence level. Subword tokenization helps large-vocabulary languages.  
*Case Study Examples:* Tokenizing SMS for spam detection; preparing tweets for analysis.

**Stop Words**  
*Definition:* Common, semantically-weak words (e.g., the, is, and) often removed from analysis.  
*When to use:* Peeprocessing to highlight semantically-meaningful content.  
*Detailed Explanation:* Removing stop words speeds up and may improve many statistical NLP models.  
*Case Study Examples:* Feature selection in sentiment analysis.

**6. Bag of Words, TF-IDF, N-grams**

**Bag of Words (BOW)**  
*Definition:* Binary or count-based representation of text ignoring word order.  
*When to use:* Simple baseline for most text classification.  
*Detailed Explanation:* Only captures word frequency; context is ignored; easy to implement and interpret.  
*Case Study Examples:* News article classification; spam detection.

**TF-IDF**  
*Definition:* Statistical measure weighting words higher if frequent in a document but rare in the corpus.  
*When to use:* Informative word detection or document similarity.  
*Detailed Explanation:* Words common to all documents have low scores; unique or rare words are weighted most.  
*Case Study Examples:* Keyword extraction; clustering documents.

**N-grams**  
*Definition:* Continuous word sequences of length n.  
*When to use:* Capturing local context or phrase structure.  
*Detailed Explanation:* Bigrams capture phrase-level context. Trigrams and higher n capture longer dependencies but increase data sparsity.  
*Case Study Examples:* Phrase detection in chat logs; search engine indexing.

**7. Word2Vec, GloVe, FastText**

**Word2Vec**  
*Definition:* Neural network method learning word vector representations.  
*When to use:* Applications needing semantic similarity or word analogy tasks.  
*Detailed Explanation:* Trained on local context; similar word vectors cluster together in dense space.  
*Case Study Examples:* Word similarity for search engines; thesaurus building.

**GloVe**  
*Definition:* Embeddings based on global co-occurrence statistics.  
*When to use:* Semantic relationships for clustering or classification.  
*Detailed Explanation:* Trained on entire dataset's word co-occurrence; good for capturing global corpus structure.  
*Case Study Examples:* Named entity recognition; emotion detection.

**FastText**  
*Definition:* Enhances embeddings by using subword (character n-gram) information.  
*When to use:* Tasks requiring robustness to rare or misspelled words.  
*Detailed Explanation:* Handles morphologically rich languages and out-of-vocabulary words by composing words from character n-grams.  
*Case Study Examples:* Language translation; text categorization.

**8. Text Classification or Sentiment Analysis**

**Text Classification**  
*Definition:* Assigning predefined labels (categories) to text.  
*When to use:* Automatic organization or tagging tasks.  
*Detailed Explanation:* Classifiers trained with features (BOW, TF-IDF, embeddings) to predict topics or types.  
*Case Study Examples:* Email spam detection; news topic classification.

**Sentiment Analysis**  
*Definition:* Determining positive/negative/neutral emotion or tone in text.  
*When to use:* Customer reviews, social media, or brand analysis.  
*Detailed Explanation:* Uses class probabilities to estimate emotion or polarity.  
*Case Study Examples:* Analyzing sentiments in movie reviews; customer support tickets.

**9. Transformers, BERT, T5, GPT**

**Transformers**  
*Definition:* Network architecture using self-attention to process sequences efficiently.  
*When to use:* NLP tasks requiring contextual awareness and long-range dependencies.  
*Detailed Explanation:* Handles sequence data in parallel instead of step-wise; powers BERT, GPT, T5 models.  
*Case Study Examples:* Text summarization; machine translation.

**BERT**  
*Definition:* Bidirectional transformer pre-trained with masked language modeling for contextual understanding.  
*When to use:* Transfer learning for most NLP tasks.  
*Detailed Explanation:* Trained to predict masked words and next sentence, producing contextual embeddings for downstream fine-tuning.  
*Case Study Examples:* Question answering; text classification.

**T5**  
*Definition:* Text-to-Text Transfer Transformer converts all NLP tasks into text generation.  
*When to use:* Flexible framework for translation, summarization, and classification.  
*Detailed Explanation:* Trained on many NLP tasks in a text-to-text framework, enabling versatile downstream use.  
*Case Study Examples:* Summarizing new articles; translating code.

**GPT**  
*Definition:* Generative Transformer trained to predict next word in text, excelling at open-ended text generation.  
*When to use:* Text, story, or dialogue generation.  
*Detailed Explanation:* Learns language modeling with an autoregressive objective; famous for chatbots and creative tasks.  
*Case Study Examples:* Chatbots; automated essay writing.

**10. Few-shot, Chain-of-Thought, Zero-shot**

**Few-shot Learning**  
*Definition:* Models generalize from only a few labeled examples.  
*When to use:* Data-poor scenarios or rapid customization.  
*Detailed Explanation:* Relies on pre-trained models or meta-learning for adaptation; common in modern LLMs.  
*Case Study Examples:* Custom text classifiers; adapting sentiment analysis to niche domains.

**Chain-of-Thought**  
*Definition:* Prompting strategy guiding models to show intermediate steps for complex tasks.  
*When to use:* Tasks requiring explicit logical or multi-step reasoning.  
*Detailed Explanation:* Encourages models to articulate their process, resulting in better performance for tasks like math or QA.  
*Case Study Examples:* Solving math word problems; stepwise reasoning in science exams.

**Zero-shot Learning**  
*Definition:* Model performs tasks without any seen labeled data for that task.  
*When to use:* Cross-domain generalization or rare/unseen categories.  
*Detailed Explanation:* Employs knowledge transfer from related tasks or domains, enabling predictions on new categories.  
*Case Study Examples:* Translating rare languages; classifying unseen document types.