

Interaction between Model Based Signal and Image Processing, Machine Learning and Artificial Intelligence [†]

Ali Mohammad-Djafari ^{1,2}

¹ Laboratoire des Signaux et Systèmes, CNRS, CentraleSupélec-Univ Paris Saclay, 91192 Gif-sur-Yvette, France; djafari@free.fr

² International Science Consulting and Training (ISCT), 91440 Bures sur Yvette, France

[†] Presented at the 39th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, 30 June–5 July 2019.

Published: 28 November 2019

Abstract: Signale and image processing has always been the main tools in many area and in particular in Medical and Biomedical applications. Nowadays, there are great number of toolboxes, general purpose and very specialized, in which classical techniques are implemented and can be used: all the transformation based methods (Fourier, Wavelets, ...) as well as model based and iterative regularization methods. Statistical methods have also shown their success in some area when parametric models are available. Bayesian inference based methods had great success, in particular, when the data are noisy, uncertain, incomplete (missing values) or with outliers and where there is a need to quantify uncertainties. In some applications, nowadays, we have more and more data. To use these “Big Data” to extract more knowledge, the Machine Learning and Artificial Intelligence tools have shown success and became mandatory. However, even if in many domains of Machine Learning such as classification and clustering these methods have shown success, their use in real scientific problems are limited. The main reasons are twofold: First, the users of these tools cannot explain the reasons when the are successful and when they are not. The second is that, in general, these tools can not quantify the remaining uncertainties. Model based and Bayesian inference approach have been very successful in linear inverse problems. However, adjusting the hyper parameters is complex and the cost of the computation is high. The Convolutional Neural Networks (CNN) and Deep Learning (DL) tools can be useful for pushing farther these limits. At the other side, the Model based methods can be helpful for the selection of the structure of CNN and DL which are crucial in ML success. In this work, I first provide an overview and then a survey of the aforementioned methods and explore the possible interactions between them.

Keywords: signal and image processing; transform based; model based; regularization; Bayesian inference; Gauss-Markov-Potts; variational Bayesian approach; Machine Learning; Artificial Intelligence

1. Introduction

Nowadays, there are great number of general purpose and very specialized toolboxes, in which, classical and advanced techniques of signal and image processing methods are implemented and can be used. Between them, we can mention all the transformation based methods (Fourier, Hilbert, Wavelets, Radon, Abel, ... and much more) as well as all the Model Based and iterative regularization methods. Statistical methods have also shown their success in some areas when parametric models are available.

Bayesian inference based methods had great success, in particular, when the data are noisy, uncertain, some missing and some outliers and where there is a need to account and to quantify uncertainties.

Nowadays, we have more and more data. To use these “Big Data” to extract more knowledge, the Machine Learning and Artificial Intelligence tools have shown success and became mandatory. However, even if in many domains of Machine Learning such as classification and clustering these methods have shown success, their use in real scientific problems are limited. The main reasons are twofold: First, the users of these tools can not explain the reasons when they are successful and when they are not. The second is that, in general, these tools can not quantify the remaining uncertainties.

Model based and Bayesian inference approach have been very successful in linear inverse problems. However, adjusting the hyper parameters is complex and the cost of the computation is high. The Convolutional Neural Networks (CNN) and Deep Learning (DL) tools can be useful for pushing farther these limits. At the other side, the Model based methods can be helpful for the selection of the structure of CNN and DL which are crucial in ML success. In this work, first I give an overview and a survey of the aforementioned methods and explore the possible interactions between them.

The rest of the paper is organized as follows: First a classification of signal and image processing methods is proposed. Then, very briefly, the Machine Learning tools are introduced. Then, through the problem of Imaging inside the body, we see the different steps from acquisition of the data, reconstruction, post-processing such as segmentation and finally the decision and conclusion of the user are presented. After mentioning some successful case studies in which the ML tools have been successful, we arrive at the main part of this paper: Looking for the possible interactions between Model based and Machine Learning tools. Finally, we mention the Open problems and challenges in both classical, model based and the ML tool.

2. Classification of Signal and Image Processing Methods

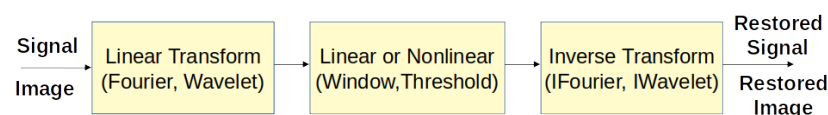
Signal and image processing methods can be classified in the following categories:

- Transform based methods
- Model based and inverse problem approach
- Regularisation methods
- Bayesian inference methods

In the first category, the main idea is to use different ways the signal and images can be represented in time, frequency, space, spacial frequency, time-frequency, wavelets, etc.

3. Transform Domain Methods

Figure 1 shows the main idea behind the transform based methods. Mainly, first a linear transform (Fourier, Wavelet, Radon, etc.) is applied to the signal or the image, then some thresholding or windowing is applied in this transform domain and finally an inverse transform is applied to obtain the result. Appropriate choices of the transform and the threshold or the widow size and shape are important for the success of such methods [1,2].



$$\begin{aligned}
 g(t) &\longrightarrow [W] \longrightarrow [T] \longrightarrow [W^{-1}] \longrightarrow f(t) \\
 g(x, y) &\longrightarrow [W] \longrightarrow [T] \longrightarrow [W^{-1}] \longrightarrow f(x, y)
 \end{aligned}$$

Figure 1. Transform methods.

4. Model Based and Inverse Problem Approach

The model based methods are related to the notions of forward model and inverse problems approach. Figure 2 shows the main idea:

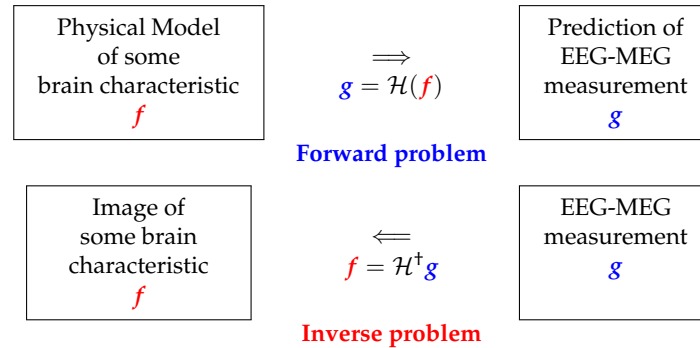


Figure 2. Model based methods.

Given the forward model \mathcal{H} and the sources f , the prediction of the data g can be done, either in a deterministic way: $g = \mathcal{H}(f)$ or via a probabilistic model: $p(g|f, \mathcal{H})$.

In the same way, given the forward model \mathcal{H} and the data g , the estimation of the unknown sources f can be done either via a deterministic method or probabilistic one. One of the deterministic method is the Generalized inversion: $f = \mathcal{H}^+(g)$. A more general method is the regularization: $\hat{f} = \arg \min_f \{J(f)\}$ with $J(f) = \|g - \mathcal{H}(f)\|^2 + \lambda \mathcal{R}(f)$ [3].

As we will see later, the only probabilistic method which can be efficiently use for the inverse problems is the Bayesian approach.

5. Regularization Methods

Let consider the linear inverse problem:

$$g = Hf + \epsilon, \quad (1)$$

Then the basic idea in regularization is to define a regularization criterion:

$$J(f) = \frac{1}{2} \|g - Hf\|_2^2 + \lambda R(f) \quad (2)$$

and optimize it to obtain the solution [4]. The first main issue in such regularization method is the choice of the regularizer. The most common examples are:

$$R(f) = \left\{ \|f\|_2^2, \|f\|_\beta^\beta, \|Df\|_2^2, \|Df\|_\beta^\beta, \sum_j \phi([Df]_j) \right\}, 1 \leq \beta \leq 2 \quad (3)$$

The second main issue in regularization is the the choice of appropriate optimization algorithm. Mainly, depending on the type of the criterion, we have:

- $R(f)$ quadratic: Gradient based, Conjugate Gradient algorithms are appropriate.
- $R(f)$ non quadratic, but convex and differentiable: Here too the Gradient based and Conjugate Gradient (CG) methods can be used, but there are also great number of convex criterion optimization algorithms.
- $R(f)$ convex but non-differentiable: Here, the notion of sub-gradient is used.

Specific cases are:

- L2 or quadratic: $J(\mathbf{f}) = \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{D}\mathbf{f}\|_2^2$.

In this case we have an analytic solution: $\hat{\mathbf{f}} = (\mathbf{H}'\mathbf{H} + \lambda\mathbf{D}'\mathbf{D})^{-1}\mathbf{H}'\mathbf{g}$. However, in practice this analytic solution is not usable in high dimensional problems. In general, as the gradient $\nabla J(\mathbf{f}) = -\mathbf{H}'(\mathbf{g} - \mathbf{H}\mathbf{f}) + 2\lambda\mathbf{D}'\mathbf{D}\mathbf{f}$ can be evaluated analytically, gradient based algorithms are used.

- L1 (TV): convex but not differentiable at zero: $J(\mathbf{f}) = \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{D}\mathbf{f}\|_1$.

The algorithms in this case use the notions of Fenchel conjugate, Dual problem, sub gradient, proximal operator, ...

- Variable splitting and Augmented Lagrangian

$$(\mathbf{f}, \mathbf{z}) = \arg \min_{\mathbf{f}, \mathbf{z}} \left\{ \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{z}\|_1 + q \|\mathbf{z}\|_2^2 \right\} \quad \text{s.t. } \mathbf{z} = \mathbf{D}\mathbf{f} \quad (4)$$

A great number of optimization algorithms have been proposed: ADMM, ISTA, FISTA, etc. [5–7].

Main limitations of deterministic regularization methods are:

- Limited choice of the regularization term. Mainly, we have: a) Smoothness (Tikhonov), b) Sparsity, Piecewise continuous (Total Variation).
- Determination of the regularization parameter. Even if there are some classical methods such as L-Curve and Cross validation, there are still controversial discussions about this.
- Quantification of the uncertainties: This is the main limitation of the deterministic methods, in particular in medical and biological applications where this point is important.

The best possible solution to push further all these limits is the Bayesian approach which has: (a) Many possibilities to choose prior models, (b) possibility of the estimation of the hyper-parameters, and most important (c) accounting for the uncertainties.

6. Bayesian Inference Methods

The simple case of the Bayes rule is:

$$p(\mathbf{f}|\mathbf{g}, \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \mathcal{M}) p(\mathbf{f}|\mathcal{M})}{p(\mathbf{g}|\mathcal{M})} \quad \text{where } p(\mathbf{g}|\mathcal{M}) = \int p(\mathbf{g}|\mathbf{f}, \mathcal{M}) p(\mathbf{f}|\mathcal{M}) d\mathbf{f} \quad (5)$$

When there are some hyper parameters which have also to be estimated, we have:

$$p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}, \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}, \mathcal{M}) p(\mathbf{f}|\boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathbf{g}|\mathcal{M})} \quad \text{where } p(\mathbf{g}|\mathcal{M}) = \int p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}, \mathcal{M}) p(\mathbf{f}|\boldsymbol{\theta}, \mathcal{M}) d\boldsymbol{\theta} d\mathbf{f} \quad (6)$$

From that joint posterior distribution, we may also obtain the marginals:

$$p(\mathbf{f}|\mathbf{g}, \mathcal{M}) = \int p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}, \mathcal{M}) d\boldsymbol{\theta} \quad \text{and} \quad p(\boldsymbol{\theta}|\mathbf{g}, \mathcal{M}) = \int p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}, \mathcal{M}) d\mathbf{f} \quad (7)$$

To be more specific, let consider the case of linear inverse problems $\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$. Then, assuming Gaussian noise, we have:

$$p(\mathbf{g}|\mathbf{f}) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, v_e \mathbf{I}) \propto \exp \left[\frac{-1}{2v_e} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 \right] \quad (8)$$

Assuming a Gaussian prior:

$$p(\mathbf{f}) \propto \exp \left[\frac{-1}{2v_f} \|\mathbf{f}\|_2^2 \right] \quad \text{or} \quad \exp \left[\frac{-1}{2v_f} \|\mathbf{D}\mathbf{f}\|_2^2 \right], \quad (9)$$

Then, we see that the posterior is also Gaussian and the MAP and Posterior Mean (PM) estimates become the same and can be computed as the minimizer of : $J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda R(\mathbf{f})$:

$$p(\mathbf{f}|\mathbf{g}) \propto \exp \left[\frac{-1}{2v_\epsilon} J(\mathbf{f}) \right] \rightarrow \hat{\mathbf{f}}_{MAP} = \arg \max_{\mathbf{f}} \{p(\mathbf{f}|\mathbf{g})\} = \arg \min_{\mathbf{f}} \{J(\mathbf{f})\} \quad (10)$$

In summary, we have:

$$\begin{cases} p(\mathbf{g}|\mathbf{f}) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, v_\epsilon \mathbf{I}) \\ p(\mathbf{f}) = \mathcal{N}(\mathbf{f}|0, v_f \mathbf{I}) \end{cases} \rightarrow \begin{cases} p(\mathbf{f}|\mathbf{g}) = \mathcal{N}(\mathbf{f}|\hat{\mathbf{f}}, \hat{\Sigma}) \\ \hat{\mathbf{f}} = [\mathbf{H}'\mathbf{H} + \lambda \mathbf{I}]^{-1} \mathbf{H}'\mathbf{g} \\ \hat{\Sigma} = v_\epsilon [\mathbf{H}'\mathbf{H} + \lambda \mathbf{I}]^{-1}, \quad \lambda = \frac{v_\epsilon}{v_f} \end{cases} \quad (11)$$

For the case where the hyper parameters v_ϵ and v_f are unknown (Unsupervised case), we can derive the following:

$$\begin{cases} p(\mathbf{g}|\mathbf{f}, v_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, v_\epsilon \mathbf{I}) \\ p(\mathbf{f}|v_f) = \mathcal{N}(\mathbf{f}|0, v_f \mathbf{I}) \\ p(v_\epsilon) = \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \\ p(v_f) = \mathcal{IG}(v_f|\alpha_{f_0}, \beta_{f_0}) \end{cases} \rightarrow \begin{cases} p(\mathbf{f}|\mathbf{g}, v_\epsilon, v_f) = \mathcal{N}(\mathbf{f}|\hat{\mathbf{f}}, \hat{\Sigma}) \\ \hat{\mathbf{f}} = [\mathbf{H}'\mathbf{H} + \hat{\lambda} \mathbf{I}]^{-1} \mathbf{H}'\mathbf{g} \\ \hat{\Sigma} = \hat{v}_\epsilon [\mathbf{H}'\mathbf{H} + \hat{\lambda} \mathbf{I}]^{-1}, \quad \hat{\lambda} = \frac{\hat{v}_\epsilon}{\hat{v}_f} \\ p(v_\epsilon|\mathbf{g}, \mathbf{f}) = \mathcal{IG}(v_\epsilon|\tilde{\alpha}_\epsilon, \tilde{\beta}_\epsilon) \\ p(v_f|\mathbf{g}, \mathbf{f}) = \mathcal{IG}(v_f|\tilde{\alpha}_f, \tilde{\beta}_f) \\ \tilde{\alpha}_\epsilon, \tilde{\beta}_\epsilon, \tilde{\alpha}_f, \tilde{\beta}_f \end{cases} \quad (12)$$

where the expressions for $\tilde{\alpha}_\epsilon, \tilde{\beta}_\epsilon, \tilde{\alpha}_f, \tilde{\beta}_f$ can be found in [8].

The joint posterior can be written as:

$$p(\mathbf{f}, v_\epsilon, v_f|\mathbf{g}) \propto \exp [-J(\mathbf{f}, v_\epsilon, v_f)] \quad (13)$$

From this expression, we have different expansion possibilities:

- JMAP: Alternate optimization with respect to $\mathbf{f}, v_\epsilon, v_f$:

$$J(\mathbf{f}, v_\epsilon, v_f) = \frac{1}{2v_\epsilon} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \frac{1}{2v_f} \|\mathbf{f}\|_2^2 + (\alpha_{\epsilon_0} + 1) \ln v_\epsilon + \frac{\beta_{\epsilon_0}}{v_\epsilon} + (\alpha_{f_0} + 1) \ln v_f + \frac{\beta_{f_0}}{v_f} \quad (14)$$

- Gibbs sampling MCMC:

$$\mathbf{f} \sim p(\mathbf{f}, v_\epsilon, v_f|\mathbf{g}) \rightarrow v_\epsilon \sim p(v_\epsilon|\mathbf{g}, \mathbf{f}) \rightarrow v_f \sim p(v_f|\mathbf{g}, \mathbf{f}) \quad (15)$$

- Variational Bayesian Approximation: Approximate $p(\mathbf{f}, v_\epsilon, v_f|\mathbf{g})$ by a separable one $q(\mathbf{f}, v_\epsilon, v_f) = q_1(\mathbf{f})q_2(v_\epsilon)q_3(v_f)$ minimizing $\text{KL}(q|p)$ [8].

7. Imaging inside the Body: From Acquisition to Decision

To introduce the link between the different model based methods and the Machine Learning tools, let consider the case of medical imaging, from the acquisition to the decision steps:

- Data acquisition:

$$\text{Object } \mathbf{f} \rightarrow \boxed{\text{Scanner}} \rightarrow \text{Data } \mathbf{g}$$

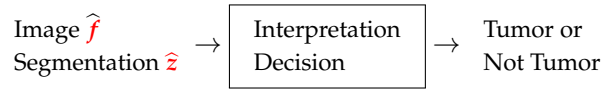
- Reconstruction:

$$\text{Data } \mathbf{g} \rightarrow \boxed{\text{Reconstruction}} \rightarrow \text{Image } \hat{\mathbf{f}}$$

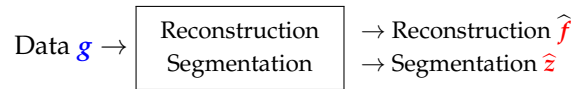
- Post Processing (Segmentation):

$$\text{Image } \hat{\mathbf{f}} \rightarrow \boxed{\text{Segmentation}} \rightarrow \hat{\mathbf{z}}$$

- Understanding and Decision:



The questions now are: Can we join any of these steps? Can we go directly from the image to the decision? For the first one, the Bayesian approach can provide a solution:



The main tool here is to introduce a hidden variable which can represent the segmentation. A solution is to introduce a classification hidden variable z with $z_j = \{1, 2, \dots, K\}$. Then in Figure 3, we have in summary:

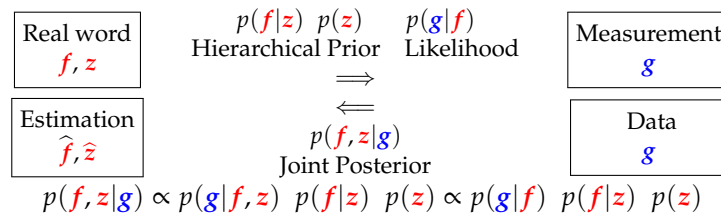


Figure 3. Bayesian approach for joint reconstruction and segmentation.

A few comments for these relations:

- $p(g|f, z)$ does not depend on z , so it can be written as $p(g|f)$.
- We may choose a Markovian Potts model for $p(z)$ to obtain more compact homogeneous regions [8,9].
- If we choose for $p(f|z)$ a Gaussian law, then $p(f, z|g)$ becomes a Gauss-Markov-Potts model [8].
- We can use the joint posterior $p(f, z|g)$ to infer on (f, z) : We may just do JMAP: $(\hat{f}, \hat{z}) = \arg \max \{p(f, z|g)\}$ or trying to access to the expected posterior values by using the Variational Bayesian Approximation (VBA) techniques [8,10–13].

This scheme can be extended to consider the estimation of the hyper parameters too. Figure 4 shows this.

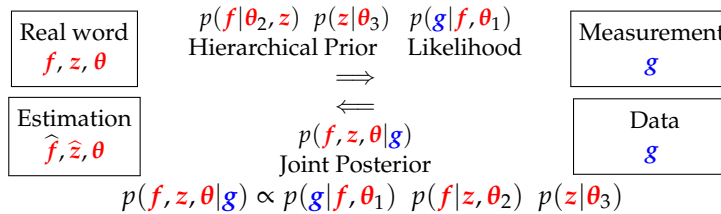


Figure 4. Advanced Bayesian approach for joint reconstruction and segmentation.

Again, here, we can use the joint posterior $p(f, z, \theta|g)$ to infer on all the unknowns [12].

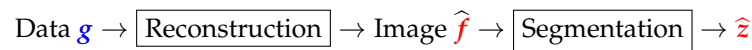
8. Advantages of the Bayesian Framework

- Large flexibility of Prior models prior
 - Smoothness (Gaussian, Gauss-Markov)

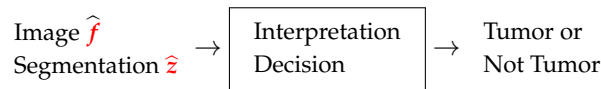
- Direct Sparsity (Double Exp, Heavy-tailed distributions)
- Sparsity in the Transform domain (Double Exp, Heavy-tailed distributions on the WT coefficients)
- Piecewise continuous (DE or Student-t on the gradient)
- Objects composed of only a few materials (Gauss-Markov-Potts), ...
- Possibility of estimating hyper-parameters via JMAP or VBA
- Natural ways to take account for uncertainties and quantify the remaining uncertainties.

9. Imaging inside the Body for Tumor Detection

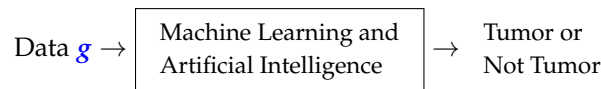
- Reconstruction and Segmentation



- Understanding and Decision



- Can we do all together in a more easily way?
- Machine Learning and Artificial Intelligence tools may propose solutions

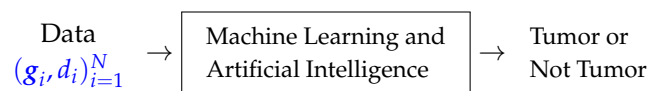


- **Learning from a great number of data**

From here, we may just do JMAP: $(\hat{f}, \hat{z}, \hat{\theta}) = \arg \max \{p(f, z, \theta | g)\}$, or better, use the Variational Bayesian Approximation (VBA) to do inference.

10. Machine Learning Basic Idea

The main idea in Machine Learning is to learn from a great number of data: $(g_i, d_i), i = 1, \dots, N$:



Between the basic tasks we can mention: (a) Classification (Tumor/Not Tumor), (b) Regression (Continuous parameter estimation) and (c) Clustering when the data have not yet labels.

Between the existing ML tools we may mention: Support Vector Machines (SVM), Decision-Tree learning (DT), Artificial Neural Networks (ANN), Bayesian Networks (BN), HMM and Random Forest (RF), Mixture Models (GMM, SMM, ...), KNN, Kmeans,...

The frontiers between Image processing, Computer vision, Machine Learning & Artificial intelligence are not very precise as it is shown in Figure 5.

Image Processing	Acquisition, Representation, Compression, Transmission
Computer Vision	Enhancement, Restoration Segmentation
Machine Learning	Edge/Feature extraction Pattern matching Objects detection
Artificial Intelligence	Understanding, Recognition, 3D

Figure 5. Frontiers between Image processing (IP), Computer vision (CV), Machine Learning (ML) and Artificial intelligence (AI).

Between the Machine learning tools, we may mention Neural Networks (NN), Artificial NN (ANN), Convolutional NN (CNN), Recurrent NN (RNN), Deep Learning (DL). This last one had shown great success in Speech Recognition, Computer Vision and specifically in Segmentation, Classification and Clustering and in Multi-modality and cross-domain information fusion [14–17]. However, there are still many limitations: Lack of interpretability, reliability and uncertainty and No reasoning and explaining capabilities [18]. To overcome, there still much to do with the Fundamentals.

11. Interaction between Model Based and Machine Learning Tools

To show the possibilities of the interaction between classical and machine learning, let consider a few examples. The first one is the case of linear inverse problems and quadratic regularization or the Bayesian with Gaussian priors. The solution has an analytic expression:

$$g = Hf + \epsilon \rightarrow \hat{f} = (HH^t + \lambda DD^t)^{-1} H^t g = BH^t g$$

which can be presented schematically as

$$g \rightarrow \boxed{H^t} \rightarrow \boxed{B} \rightarrow \hat{f} \text{ or directly } g \rightarrow \boxed{\text{CNN or DL}} \rightarrow \hat{f}$$

As we can see, this induces directly a linear NN structure. In particular, if H represents a convolution operator, then H^t and $H^t H$ are too and probably the operator B can also be well approximated by a convolution and the whole inversion can be modelled by a CNN [19].

The second example is the denoising $g = f + \epsilon$ with L1 regularizer, or equivalently, the MAP estimator with a double exponential prior, where the solution can be obtained by a convolution followed by a thresholding [20,21].

$$g = f + \epsilon \rightarrow \hat{f} = \frac{1}{\lambda} H^t g \text{ followed by Thresholding}$$

$$g \rightarrow \boxed{H^t} \rightarrow \boxed{\text{Thresholding}} \rightarrow \hat{f} \text{ or directly } g \rightarrow \boxed{\text{CNN or DL}} \rightarrow \hat{f}$$

The third example is the Joint Reconstruction and Segmentation that was presented in previous sections. If we present the different steps of reconstruction, segmentation and parameter estimation, we can also compare it with some kind of NN (combination of CNN, RNN and GAN) [22,23].

$$g \rightarrow \boxed{\text{Reconstruction}} \rightarrow \hat{f} \rightarrow \boxed{\text{Segmentation}} \rightarrow \hat{z} \rightarrow \boxed{\text{Parameter estimation}} \rightarrow \hat{\theta}$$

12. Conclusions and Challenges

Signal and image processing, imaging systems and computer vision have made great progress in the last forty years. The first category of the methods was based on linear transformation followed by

a thresholding or windowing and coming back. The second generation was model based: forward modeling and inverse problems approach. The main successful approach was based on regularization methods using a combined criterion. The third generation was model based but probabilistic and using the Bayes rule, the so called Bayesian approach. Nowadays, Machine Learning (ML), Neural Networks (NN), Convolutional NN (CNN), Deep Learning (DL) and Artificial Intelligence (AI) methods have obtained great success in classification, clustering, object detection, speech and face recognition, etc. But, they need great number of training data and lack still explanation and they may fail very easily. For signal and image processing or inverse problems, they need still progress. This progress is coming via their interaction with the model based methods. In fact, the successful of CNN and DL methods greatly depends on the appropriate choice of the network structure. This choice can be guided by the model based methods. For inverse problems, when the forward models are not available or too complex, NN and DL may be helpfull. However, we may still need to choose the structure of the NN via approximate forward model and approximate Bayesian inversion.

References

1. Carre, P.; Andres, E. Discrete analytical ridgelet transform. *Signal Process.* **2004**, *84*, 2165–2173.
2. Dettori, L.; Semler, L. A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography. *Comput. Biol. Med.* **2007**, *37*, 486–498.
3. Mohammad-Djafari, A. *Inverse Problems in Vision and 3D Tomography*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
4. Idier, J. *Bayesian Approach to Inverse Problems*; John Wiley & Sons: New York, NY, USA, 2008.
5. Afonso, M.V.; Bioucas-Dias, J.M.; Figueiredo, M.A. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.* **2010**, *19*, 2345–2356.
6. Bioucas-Dias, J.M.; Figueiredo, M.A. An iterative algorithm for linear inverse problems with compound regularizers. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 685–688.
7. Florea, M.I.; Vorobyov, S.A. A robust FISTA-like algorithm. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 4521–4525.
8. Chapdelaine, C.; Mohammad-Djafari, A.; Gac, N.; Parra, E. A 3D Bayesian Computed Tomography Reconstruction Algorithm with Gauss-Markov-Potts Prior Model and its Application to Real Data. *Fundam. Informaticae* **2017**, *155*, 373–405.
9. Féron, O.; Duchêne, B.; Mohammad-Djafari, A. Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data. *Inverse Probl.* **2005**, *21*, S95.
10. Chapdelaine, C.; Mohammad-Djafari, A.; Gac, N.; Parra, E. A Joint Segmentation and Reconstruction Algorithm for 3D Bayesian Computed Tomography Using Gauss-Markov-Potts Prior Model. In Proceedings of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017), New Orleans, LA, USA, 5–9 March 2017.
11. Chapdelaine, C.; Gac, N.; Mohammad-Djafari, A.; Parra, E. New GPU implementation of Separable Footprint Projector and Backprojector: First results. In Proceedings of the 5th International Conference on Image Formation in X-ray Computed Tomography, Salt Lake City, UT, USA, 20–23 May 2018.
12. Ayasso, H.; Mohammad-Djafari, A. Joint NDT image restoration and segmentation using Gauss–Markov–Potts prior models and variational bayesian computation. *IEEE Trans. Image Process.* **2010**, *19*, 2265–2277.
13. Chapdelaine, C. Variational Bayesian Approach and Gauss-Markov-Potts prior model. *arXiv* **2018**, arXiv:1808.09552.
14. Szegedy, C.; Toshev, A.; Erhan, D. Deep Neural Networks for Object Detection. In Proceedings of the Advances in Neural Information Processing Systems 26 (NIPS 2013), Lake Tahoe, NV, USA, 5–10 December 2013; pp. 2553–2561.
15. Deng, L.; Yu, D. *Deep Learning: Methods and Applications*; NOW Publishers: Hanover, MA, USA, 2014.
16. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117.

17. Schmidhuber, J. The New AI: General & Sound & Relevant for Physics. In *Artificial General Intelligence*; Goertzel, B., Pennachin, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 175–198.
18. Chellappa, R.; Fukushima, K.; Katsaggelos, A.; Kung, S.Y.; LeCun, Y.; Nasrabadi, N.M.; Poggio, T.A. Applications of Artificial Neural Networks to Image Processing (guest editorial). *IEEE Trans. Image Process.* **1998**, *7*, 1093–1097.
19. Ciresan, D.C.; Meier, U.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. Flexible, High Performance Convolutional Neural Networks for Image Classification. In Proceedings of the Twenty-Second International joint conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; pp. 1237–1242.
20. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012; p. 4.
21. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. *arXiv* **2013**, arXiv:1311.2901.
22. Masci, J.; Meier, U.; Ciresan, D.; Schmidhuber, J. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. In *Artificial Neural Networks and Machine Learning—ICANN 2011*; Honkela, T., Duch, W., Girolami, M.A., Kaski, S., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6791, pp. 52–59.
23. LeCun, Y.; Bengio, Y. Convolutional Networks for Images, Speech, and Time-Series. In *The Handbook of Brain Theory and Neural Networks*; Arbib, M.A., Ed.; MIT Press: Cambridge, MA, USA, 1995.



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).